BIOCHEMICAL STUDY OF A PSI-LHCI COMPLEX AND MOLECULAR STUDY OF *fcp* GENES IN THE DIATOM *CYLINDROTHECA FUSIFORMIS*

By

Hua Zhang

RECOMMENDED:

Advisory Committee Chair

Department Head

APPROVED:

Dean, College of Science, Engineering & Mathematics

Dean of the Graduate School

8-12-02

Date

BIOCHEMICAL STUDY OF A PSI-LHCI COMPLEX AND MOLECULAR STUDY OF *fcp* GENES IN THE DIATOM *CYLINDROTHECA FUSIFORMIS*

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Hua Zhang, B.S.

Fairbanks, Alaska

August 2002

ABSTRACT

A photosystem I-light-harvesting complex I (PSI-LHCI) was isolated from oxygen-evolving thylakoids of the diatom *Cylindrotheca fusiformis*. The circular dichroism (CD) spectrum of the complex resembled the photosystem I (PSI) complex of green plants. A single 77K fluorescence emission was observed at 715 nm. The excitation spectrum confirmed that both chlorophyll *c* and carotenoids were energetically coupled to chlorophyll *a*. The complex contained PSI reaction center proteins PsaA/B, the PSI accessory subunit PsaC, and nine light-harvesting complex (LHC) apoproteins including an 18kD and a 17.5kD protein. Photosystem II core polypeptides were not detected by immuno- or silver staining. Taken together, the CD, fluorescence, and protein data indicate that at least nine LHC apoproteins can be specifically associated with PSI in this diatom. Twenty *fcp* gene sequences that encode fucoxanthin-chlorophyll *a*/*c* light-harvesting proteins (FCPs) were obtained. Eight of them encoded an abundant 18kD FCP and three encoded proteins that are similar to a 17.5kD FCP, but it was not possible to conclusively confirm that any of these genes encode proteins associated with LHCI.

Table of Contents

List of Figures

List of Tables

Acknowledgement

First of all, I would like to thank my major advisor, Dr. Gerald Plumley, for his guidance, encouragement, and great help throughout my graduate study. I also wish to thank my committee members, Dr. Bert Boyer and Dr. Larry Duffy, for their helpful advice and support.

I am grateful to many people who have helped and encouraged me during my study here. Dr. Tracey Martinson gave me hands-on training and many useful suggestions for my protein work (Chapter 2); Sheila Chapin was a constant source of guidance from the first day I became a graduate student at UAF; Julie Matweyou, Carla Richardson, and Jayme Ballentyne were always there when I needed them.

Special thanks go to my family in China, and my husband, Xiaoming Zhang, for their love and support.

The Department of Chemistry and Biochemistry, Alaska NSF EPSCoR, NSF RIDGE, NASA Astrobiology, and NASA Astrobiology through ASU provided financial support.

Chapter 1 General Introduction

Photosynthesis

Photosynthesis is the process by which plants use light energy to convert CO_2 and H_2O to carbohydrate and O_2. Light energy is absorbed by protein-bound chlorophylls and carotenoids that are located in the light-harvesting complexes (LHCs). The absorbed energy migrates to photosynthetic reaction centers (RCs) where the reaction center chlorophylls are located. These special chlorophyll molecules are able to transfer their high-energy electrons to the photosynthetic electron transport chain. Electrons derived from water oxidation are used to replace those lost from the RC. In photosynthesis, light energy is converted to chemical energy that is stored as carbohydrate. It is estimated that photosynthesis fixes ~ 10^{11} tons of carbon annually, which represents the storage of over 10^{18} KJ of energy. Approximately 50% of this production occurs in the world's oceans, mostly at high latitudes including Alaska's coastal regions.

There are two types of photosynthesis: oxygenic and anoxygenic. Anoxygenic photosynthesis occurs in photosynthetic bacteria, while oxygenic photosynthesis occurs in cyanobacteria, eukaryotic plants, and algae. There are four known anoxygenic photosynthetic lineages, the green sulfur bacteria, the green non-sulfur bacteria, heliobacteria, and purple bacteria. The phylogenetic relationships of these different bacteria are incredibly complicated and have led to a realization that the evolution of photosynthesis (a few genes) is distinct from the evolution of photosynthetic organisms. Horizontal gene transfer of photosynthetic genes appears to have been the norm, rather than the exception (Xiong et al. 2000). Anoxygenic photosynthesis utilizes materials other than H_2O as a source of electrons and therefore does not produce O_2. Oxygenic photosynthesis is the major type of photosynthesis on earth and it involves most photosynthetic organisms. Oxygenic photosynthesis can be summarized by the equation:

$$6\ CO_2 + 6\ H_2O + \text{light} \text{-------}> C_6H_{12}O_6 + 6\ O_2.$$

Oxygenic photosynthesis is a two-stage process: the light reactions and the dark reactions. The focus of this thesis is on the light reactions and the pigment-protein complexes involved in these reactions.

Photosystem I and photosystem II

In plants and cyanobacteria, two reaction center pigment- protein complexes are involved in photosynthesis: photosystem I and photosystem II (PSI and PSII, respectively). Each photosystem is composed of a RC Core complex that contains the reaction center chlorophylls, a set of core antenna pigment molecules, and one or more peripheral light-harvesting complexes. Within the chloroplast thylakoid membrane, the peripheral light-harvesting complexes surround the reaction centers and serve as antennae. Light energy absorbed by these complexes is funneled into the reaction centers, and as a result, drives the light-dependent photosynthetic reactions. The arrangement of two photosystems, as well as other components of the photosynthetic apparatus in thylakoid membranes, is known as the "Z-scheme"(Hill and Bendall, 1960), in reference to the "Z" appearance that early researchers sketched in their notebooks as they attempted to understand photosynthetic electron flow. Each photosystem is independently activated by light, but electrons flow either from PSII to PSI and to NADPH in noncyclic photosynthesis or around PSI during cyclic photosynthesis. PSI and PSII operate in series to couple H_2O oxidation with $NADP^+$ reduction. H_2O is oxidized to O_2 within PSII and the light-driven electrons transported from PSII to the cytochrome (Cyt) b_6f complex and then to PSI where they are used to reduce $NADP^+$. The resulting proton gradient, generated in part from the oxidation of water and in part from H^+ pumping at Cyt b_6f, powers the synthesis of ATP by the CF_1F_0 proton-translocating ATP synthase.

Earlier studies of photosystem II revealed as many as 22 subunits (products of the psbA-psbV genes; Masojidek et al., 1987) but recent studies have found additional proteins such as PsbX (Shi et al., 1999) and PsbZ (Swiatek et al., 2001). Importantly, PsbZ plays

an important role in controlling the interactions of PSII particles with the light-harvesting antenna (Swiatek et al. 2001). The RC chlorophylls are positioned at the heart of the PsbA/PsbD heterodimer (Zouni et al., 2001). The three-dimensional structures of the reaction center core complex of purple bacteria and PSII of cyanobacteria, both belonging to the so-called "type II reaction center," have been determined to 2 Å and 3.8Å, respectively (Lancaster et al., 2000; Zouni et al., 2001). A PSII sub-complex containing D1, D2, CP47, and cytochrome *b*-559 of spinach has been determined to 8Å (Rhee et al., 1998). The arrangement of pigments and proteins in this complex resembled that of purple bacteria and of plant PSI, indicating a common evolutionary origin for these assemblies.

Photosystem I (PSI) is comprised of at least 12 subunits, products of the psaA-O genes (Ikeuchi, 1992). The RC chlorophylls are on the PsaA/PsaB heterodimer (Jordan et al., 2001). Three subunits, PsaG, PsaH, and PsaN, are found in green plants, but not in the cyanobacterial PSI. The other 11 subunits of the PSI core were conserved but several proteins apparently have a different role in cyanobacteria than in green plants. The structure and function of photosystem I has been well studied in cyanobacteia and green plants (see review by Chitnis, 2001; Scheller et al., 2001; Fromme et al., 2001; Golbeck, 1994; Gobets et al., 2001). The three-dimensional structure of cyanobacterial PSI has recently been determined at 2.5 Å (Jordan et al., 2001). Twelve protein subunits with 96 chlorophylls, 22 carotenoids, three 4Fe-4S clusters and other cofactors were bound to the PsaA/B core proteins.

Core light-harvesting antenna

In plants, algae and cyanobacteria, each reaction center has an internal antenna. In PSII, the internal antennae consist of two chlorophyll *a*-binding complexes, CP47 and CP43 (Camm and Green, 1983) that are products of the psbB and psbC genes, respectively (Morris and Herrmann, 1984; Alt et al., 1984). The chlorophylls that constitute the

internal antenna of PSI are located on the PsaA and PsaB core polypeptides along with the reaction center chlorophylls (Mullet et al., 1980). Beta-carotene is also present in the core antenna complexes of PSI and PSII, with the exception of a few evolutionarily divergent algae that use a different carotene isomer. These core antenna serve as a bridge between the RC chlorophylls and the more abundant peripheral light-harvesting antenna complexes.

Peripheral light-harvesting antenna

Most of the chlorophyll in green plants and eukaryotic algae is associated with peripheral antenna light-harvesting complexes known as LHCs. The LHC proteins (LHCPs) of algae and green plants are structurally related, implying a common ancestry (Green et al., 1991). All LHCPs bind xanthophylls and chlorophylls. The LHCPs of green plants bind both chlorophyll *a* (Chl *a*) and chlorophyll *b* (Chl *b)*, and are called chlorophyll *a/b*-binding proteins (CAB) while the LHCPs of chromophytes such as diatoms bind Chl *a*, chlorophyll *c* (Chl *c*), and fucoxanthin and are called fucoxanthin-chlorophyll *a/c* binding proteins (FCP). The consensus sequence for the "generic LHC motif" is ELINGRLAMLGFLGFLVPELIT (Jansson 1999).

The LHCPs contain three distinct trans-membrane helices, with the first and the third helix sharing distinct homology. It has been proposed that these proteins arose from a two-helix ancestral polypeptide (Green and Pichersky, 1994). Gene duplication and fusion of the coding region resulted in a four-helix protein. The subsequent loss of the fourth helix resulted in the final three-helix form, though at least one LHCP-like protein in green plants, PsbS, retains four helices (Green and Kühlbrandt, 1995) and appears to play a crucial role near the PSII reaction center (Li et al., 2000). The three-dimensional structure of green plant LHCII has been determined to 3.4Å (Kühlbrandt et al., 1994). Each LHCII protein binds a minimum of 12 chlorophylls and three to four carotenoids for light harvesting and energy transfer to the RC of PSI and PSII. The LHCPs are encoded

by nuclear genes, synthesized in the cytosol and post-translationally transported into chloroplasts and inserted into the thylakoid membrane (Jansson, 1994).

In higher plants, the LHCs have been given distinct names representative of their unique functions. The five chlorophyll *a*/*b*-binding peripheral antennae are CP29, CP26, CP24, LHCI, and LHCII (Camm and Green, 1983; Bassi et al., 1987; Dunahay and Staehelin, 1986). CP29, CP26, and CP24 are minor antennae associated with PSII, and are energetically and physically connected to the PSII core antenna, CP47 and/or CP43. LHCII is the major antennae associated with PSII. A subset of the LHCII antenna is mobile and associates with either PSI or PSII depending upon the status of a phosphothreonine residue (Mullet, 1983) that is acted on by LHC kinase, an enzyme whose activity is dependent upon the redox status of the chloroplast. The PSI-specific antenna is referred to as LHCI. This complex is associated on one side of the PSI core (Boekema et al., 2001) and it was suggested that the mobile LHCII funnel their energy through LHCI rather than directly to the PSI core (Knoetzel and Simpson, 1991). However, a recent study showed that LHCII appears to dock directly with the PSI core, presumably making contact with PSI-H and possibly PSI-L (Lunde et al., 2000). In addition to these main LHC antenna, the LHC family is considered to have "extended" family members that consists of distant relatives, such as the early-light-induced-proteins (ELIPs; Green et al., 1991), high-light-induced-proteins (HLIPs; Adamska, 1997), and the PsbS-family proteins (Funk et al., 1994).

The LHCI complex

The LHCI antenna complex of green plants consists of four to six different polypeptides, encoded by the Lhca1-6 genes, with molecular weights ranging from 20 to 24kDa (Jansson, 1999). The Lhca proteins bind, on average, 8 Chl *a*, 2 Chl *b* and 2 carotenoid molecules (Ihalanen et al., 2000) and exist in approximately equal amounts (Jansson, 1994). It was originally suggested that LHCI in green plants forms two kinds of

heterodimers in solution, called LHCI-730 and LHCI-680 (Haworth et al., 1983). Lhca1 and Lhca4 have been shown to form a heterodimer (Jansson et al., 1996; Schmid et al., 1997) and exhibit an emission maximum at 730nm at 77K (Jansson, 1994). In contrast, Lhca2 and Lhca3 were found to form homodimers (Jansson et al., 1996) and have a fluorescence emission maximum at 680nm. It is not clear if the Lhca5 and Lhca6 genes are expressed and, if they are, it must be at very low levels (Jansson, 1999).

LHCs of chromophytes

The peripheral antenna of green plants and green algae have been characterized in detail but very little is known about the LHCs of diatoms and other chlorophyll *c*-containing algae (i.e., the chromophytes), let alone details about specific complexes such as LHCI. Recently, a substantial number of genes encoding FCPs (*fcp* genes) from several chromophytes including diatoms and brown algae were sequenced (Apt, et al., 1995; Bhaya and Grossman, 1993; Eppard and Rhiel, 1998; Eppard et al., 2000; Kroth-Pancic, 1995; LaRoche, et al., 1994; Smith, 1997), but the specific functions of individual protein products are not known. The *fcp* genes share a homologous primary sequence to the green plant LHCP and probably have a similar tertiary structure in the membranes (Green and Durnford, 1996). The high degree of similarity between different *fcp* genes, both between and within species, has facilitated DNA sequence work (e.g., PCR using degenerate primers), but has been a curse to biochemists seeking to unravel precise functions of individual genes.

The function of individual LHCP in green plants was ascertained through a two decade long process involving dozens of labs conducting biochemical and biophysical analyses coupled with other labs that determined DNA sequences and/or generated mutants missing specific LHCP genes. Such studies are in their infancy in non-green algae. The very earliest studies of LHCs in chromophytes attempted to isolate and characterize the most abundant proteins, in large part because of their abundance, and because in the early

1980's it was not known that the LHCP gene family was so large and multi-functional. Owens and co-workers (1986) tried to integrate whole cell biophysics and LHC function (Owens, 1986; Owens and Wold, 1986) but such studies ultimately proved to be premature as additional facts were unraveled in this lab (Ting and Owens 1993). By the late 1980's it was becoming clear that the green plant LHCP gene family was large and diverse and credence was given to a largely overlooked report by Plumley and Schmidt (1983) that there were many structurally similar LHCPs in diatoms and green plants (Bhaya and Grossman, 1993). A study that included diatoms and brown algae (Berkaloff et al. 1990) reported LHCI complexes, but the proteins were not specifically identified (e.g., by immuno-staining or N-terminal amino acid sequencing). A PSI-LHCI was also reported from a number of diatoms (Brown 1988) but only "green gels" were shown, so it was not possible to know the protein composition, or purity, of these complexes. A PSI-LHCI has been reported from another alga with Chl *c* (*Rhodomonas* sp, a cryptophyte; Bathke et al. 1999), but it was not clear that the putative LHCI apoproteins were specific for PSI. Similarly, a recent study of brown algae concluded that LHCI- and LHCII-specific complexes cannot be distinguished biochemically (DeMartino et al. 2000).

A major problem with most studies of chlorophyll-protein complexes of algae with Chl *c* is that crude membrane fractions, often derived after harsh treatment of cells to obtain cell breakage (e.g., French pressure cells operated at high pressure), are the starting material for analysis. Martinson and co-workers (Martinson 1996; Martinson et al. 1998) clearly demonstrate that use of crude membrane fractions, even those that are enriched in thylakoids, have few biochemical and biophysical properties that can be linked to *in vivo* data and/or to data obtained using gentle techniques that allow purification of oxygen evolving thylakoids. While the earlier studies of diatom LHC complexes, including those that looked specifically at PSI-LHCI complexes, provide intriguing insight into the nature of diatom LHCI apoproteins, it is unlikely that specific protein functions can be assigned from crude membrane fractions heavily contaminated with non-photosynthetic proteins,

and from Chl-protein complexes whose biophysical properties have been altered due to harsh biochemical fractionation protocols.

Spectroscopic analyses of pigment-protein complexes

LHCs and RCs are pigment-protein complexes that bind different pigments with specific function: light harvesting or photochemistry. The main pigments in green plants and eukaryotic algae include Chl *a*, Chl *b*, Chl *c*, carotenoids and xanthophylls such as fucoxanthin. Chl *b* and Chl *c* as well as carotenoids are usually called accessory pigments, and pass the light energy they absorb to Chl *a*, thus increasing the spectral range for photosynthetic active radiation (PAR). There are also a pair of special Chl *a* in RCI and RCII that absorb at different wavelengths, the Chl *a* in RCII absorbs at 680nm and is referred to as P680 while the Chl *a* in RCI absorbs at 700nm and is referred to as P700. All these pigments have very different spectroscopic properties, resulting exclusively to the fact that they are found in different protein environments. The major spectroscopy tools used in the analyses of pigment-protein complexes include absorption, fluorescence emission and excitation, and circular dichroism.

In vivo all chlorophyll pigments are present in complexes with protein. In other words, there are no free chlorophyll pigments in plants or algae. In contrast, carotenoids can be part of the photosynthetic apparatus but can also be found in other locations (e.g., cell walls) and be used for non-photosynthetic purposes. The absorption spectrum of a pigment is sensitive to its physical microenviroment, which in turn affects its absorption spectrum. Chlorophylls found in chlorophyll-protein complexes generally absorb at longer wavelengths than free pigment. The extent of this absorption shift in chlorophyll absorption provides a method of identifying different pigment-protein complexes. As mentioned above, the P680 and P700 chlorophylls absorb at unique, and surprisingly long, wavelengths. The chlorophylls in the RCII and RCI core antenna absorb at somewhat shorter wavelengths, while LHCs absorb at even shorter wavelengths. Since shorter

wavelengths of light have more energy, excitation energy flows from LHC to core antenna to reaction centers.

Fluorescence is another method that can be used to characterize pigment-protein complexes. When chlorophylls absorb a photon, they are excited to a state that has higher energy than the ground state. An excited pigment is not as stable as it is in the ground state and, if not used in photochemistry, tends to decay back to the ground state by releasing energy. One way to release energy is to emit a photon. This process is called fluorescence emission. A fluorescence emission spectrum is measured by exciting a sample with an excitation light source and passing the emitted radiation through a scanning monochromator to determine the fluorescence emissions throughout a certain wavelength range (generally 650nm - 800nm). Fluorescence spectra are generally recorded at 77K so that electron transfer is prevented, thus increasing fluorescence emissions due to the absence of photochemistry, and allowing resolution of individual pigment beds, such as those associated with PSII and PSI.

The 77K fluorescence properties of green plants have been well described (Satoh and Bulter, 1978). There are three major emissions bands, 685 and 695nm associated with CP43 and CP47, part of the PSII core antenna, and 730nm, arising from PSI or, as explained below, from LHCI. The 77K spectra of diatoms reveal peaks at 685 and 717nm that have been ascribed to PSII and PSI, respectively (Martinson et al. 1998).

The 77K fluorescence spectra of purified thylakoids generally match that of whole ells, but once thylakoids are solubilized with detergents, there are noticeable changes, as the peripheral antenna are energetically disconnected from PSI and PSII. The biggest change is that a new fluorescence emission arises at 680-683nm due to dissociation of LHCII and/or LHCI. This signal is so large that it obscures those from PSII (685 and 695nm) and PSI (730nm). Chlorophylls are released from thylakoids when high detergent

concentrations are used, and even larger signals arise at 675-677nm (Chl *a*), 648-652nm (Chl *b*) or 636-640nm (Chl *c*).

Upon resolution of pigment-protein complexes from detergent-treated thylakoids (e.g., by sucrose density gradient centrifugation or non-denaturing PAGE), the 77K fluorescence properties remain largely unchanged. In green plants, LHCII fluoresces at 680-683nm while PSII fluoresces at 695nm. The fluorescence emissions of higher plant PSI are more complex. Purified PSI, devoid of LHCI, fluoresces at 720nm whereas PSI-LHCI emits at 730nm (Mullet et al., 1980; Harworth et al., 1983). The different signals from PSI vs. PSI-LHCI are, in large part, an artifact of freezing pigment-protein complexes to 77K, and in part due to the presence of "red chlorophylls" in LHCI (Gobet and Grondelle, 2001). Purified green plant LHCI (devoid of PSI) emits at 730nm, and at 680nm upon further fractionation into two sub-antenna complexes. The *in vivo* fluorescence emissions from green plant PSI/LHCI are thought to be unique: the green plant *in vivo* signal arises from LHCI, whereas in all other organisms (including diatoms) the emission signals are thought to arise from PSI core chlorophylls. However, it is important to point out that LHCI may be responsible for PSI emissions in diatoms, a point that is addressed in Chapter 2.

Fluorescence properties of diatom pigment-protein complexes have not been as well characterized as those of green algae and higher plants. Part of the problem has been that the isolation of functional LHCII, PSI, and PSII has been difficult. However, early studies reported that the major LHC of diatoms fluoresces at 675-680nm (Friedman and Alberte, 1986; Owens and Wold, 1986) and it is unclear if these LHCs were similar to LHCII or LHCI of green plants. PSI-enriched fractions have been isolated from several diatoms with very different protein compositions and fluorescence properties (Berkaloff et al., 1990; Caron and Brown, 1987; Brown, 1988). Two PSI-containing fractions of a diatom, *Phaeodactylum triconutum*, fluoresced at 691nm and 716nm (Caron and Brown, 1987); the fraction that was more enriched in P700 mainly fluoresced at 691nm. In

contrast, 'native' PSI fraction from *Phaeodactylum tricornutum* had a single fluorescence emission maximum at about 720nm (Berkaloff et al., 1990). It is not known if the fluorescence signals of these complexes were from PSI or from contaminating PSII although the assignment of the long wavelength emission was the core itself (Berkloff et al., 1990) or the associated LHC (Brown, 1988).

Fluorescence excitation spectra, in contrast to the emission spectra described above, measure the relative efficiency with which different wavelengths of light elicit fluorescence emissions (Hipkins and Baker, 1986). In essence, fluorescence excitation spectra are similar to absorption spectra, but with one key difference. In absorption spectra, every pigment will contribute to absorption whereas in excitation spectra, only those pigments that are energetically coupled to the terminal emitter are detectable. For a single pigment in solution, the absorption spectrum of the pigment corresponds to the excitation spectrum for fluorescence. For pigment-protein complexes, not all pigments are necessarily energetically coupled, as some pigments may play a structural role in assembly of the pigment-protein complex and/or a role in dissipation of excess photons (as heat or temporary electron acceptors – topics beyond the scope of this discussion). In summary, excitation spectra provide information on the pigments that contribute to the fluorescence emission and are very useful in the study of excitation energy transfer from accessory pigments to the chlorophylls in the reaction centers.

Circular dichroism (CD) measures the difference between the absorption of left-handed and right-handed circularly polarized light and is generally used to identify pigment-pigment and/or pigment-protein interactions. The interpretation of a CD signal is very complex but a CD signal indicates molecular asymmetries. If a CD spectrum of an isolated pigment-protein complex resembles that of the complex *in vivo*, this is taken as very strong evidence that the complex has not been drastically altered during the isolation. Not all pigments in a pigment-protein complex contribute to CD signals, but thus far, all pigment-protein complexes isolated have demonstrable CD signals arising from a subset

of pigments. The dichroic pigments can be detected because the amplitude of their CD signal is substantially enhanced relative to that of free pigment. In other words, pigment-protein interactions enhance CD signals arising from pigment-pigment interactions within the complex and these signals are greater than those arising from pigment not associated with protein. A commonly used "control" for CD is to heat the complex, thus dissociating pigment-protein interactions, and repeating the CD spectral scan. A reduction in signal upon heating is a clear indication that the isolated pigment-protein complex is close to its native condition and that it was not perturbed during isolation.

A final caveat about absorption, fluorescence, and CD spectroscopy relates to biochemical studies that employ detergents to solubilize pigment-protein complexes. Many different detergents have been used and it is widely accepted that each detergent affects pigments differently, presumably due to the fact that each detergent binds to different regions of the apoproteins and/or binds with different affinity. It is not unusual for two investigators to work with the same species and obtain markedly different spectra, a problem that can be "easily" traced to use of different detergents for solubilization.

Photosynthetic membranes

In photosynthetic organisms, major functional assemblies for photosynthesis are located within a membrane system. Although similarities exist in how the functional components are integrated within the membrane, the morphological arrangements of the membranes within photosynthetic cells are highly variable.

In prokaryotic green bacteria, the energy-transporting activities are contained within specialized regions of cell membranes. Light-harvesting complexes (LH) are contained in nonmembranous chlorosomes that are bound to the cytoplasmic surface of these specialized regions and the reaction centers are contained entirely within the cytoplasmic membrane (Sprague and Varga, 1986). In purple bacteria, highly differentiated

intracytoplasmic membranes with a high pigment content called chromatophores are the locations for photosynthesis. The RC and the LH are contained in a system of internal membrane called intracytoplasmic membranes (ICM's). These intracytoplasmic membranes are continuous with the cell membrane and often appear as vesicles (Sprague and Varga, 1986).

Cyanobacteria are the most primitive cells that contain thylakoids. Thylakoids in cyanobacteria are not compartmentalized within an organelle and pigment complexes called phycobilisomes are attached to the surface of the thylakoids. In higher plants and all eukaryotic algae, thylakoids are enclosed in chloroplasts that are surrounded by two to four membrane envelopes.

The arrangements of the thylakoid membranes within the chloroplast vary among photosynthetic organisms. Red algae have the simplest arrangement of thylakoids within chloroplasts – they exist individually and lie in parallel with each other. Phycobilisomes are attached to the outer surface of the membrane and there are few points of contact between adjacent membranes. Recent research (Wolfe et al., 1994) showed that a PSI associated LHCI coexists in red algae with phycobilisomes, a finding of great importance relative to the work reported here.

The thylakoids of higher plants and green algae are developed into grana lamellae that are stacks of flattened disk-shaped inter-connected membranous sacs, and stroma lamellae that are loosely arranged membranes connecting grana lamellae. The grana lamellae are predominated by PSII O_2-evolving reaction complexes and the stroma lamellae by PSI particles and ATP synthase. In chromophytes, the thylakoids are arranged in groups of three and are not segregated into granal and stromal lamellae (Gibbs, 1970) and LHC is equally distributed all along the length of the thylakoids while PSI is present on all the membranes yet more concentrated in the external membranes (Lichtlé et al., 1992).

Evolution of reaction centers, light-harvesting complexes, and photosynthesis

There are two types of reaction centers in photosynthetic organisms: Type I and Type II (Barber and Anderson 1994; Nitschke and Rutherford 1991; Goldbeck 1993; Blankenship and Hartman 1998). These are also referred to as the iron-sulfur type (Type I) and quinone-type (Type II) indicating the primary electron acceptor. These two types of reaction centers are assumed to have arisen from a common ancestral gene product in bacteria (Figure 1.1) that diverged substantially through evolutionary time to yield RCs with strikingly unique biochemical and biophysical properties in different bacterial lineages. Contemporary purple bacteria have a Type II RC while the Type I RC is found in green sulfur bacteria. The oxidative range of the purple bacterial Type II RC spans from –0.75 to 0.4 volts while the Type I RC found in green sulfur bacteria spans from -1.0 to 0.15 volts (Nitschke and Rutherford, 1991). Neither type of RC is capable of utilizing water as electron donor, and hence both are anoxygenic. It is thought that RCI was the ancestral photosystem and RCII is a descendent from RCI via gene duplication and gene splitting (Baymann et al., 2001).

An interesting hypothesis, well supported by molecular data, suggests a fusion event between a bacterium with a Type I RC and another bacterium with a Type II RC by lateral gene transfer (Blankenship, 1992), putatively giving rise to a cell with two types of RC and hence able to span a broad range of photosynthetic redox potentials. Despite having two types of RCs, the resulting progeny were unable to utilize water as electron donor. Hydrogen peroxide may have been the transitional electron donor to the Type II reaction center, setting the stage for the eventual evolution of the PSII reaction center from this ancestral protein complex (Blankenship and Hartman 1998). Once water oxidation became possible, there was strong selective pressure to link the electrons from this abundant and cheap donor to reduction of NADPH through the co-evolution of PSI from the Type I RC, resulting in a primitive organism with an oxygenic form of photosynthesis. Cyanobacteria are today considered the most primitive extant oxygenic

phototrophs and are the likely progenitor of chloroplasts through the process of primary, and in some cases, secondary endosymbiotic events (Figure 1.1).

Different lineages of anoxygenic bacteria evolved unique light-harvesting mechanisms to supplement photon capture and electron transfer by the RC core. For instance, purple bacteria antenna complexes consist of light-harvesting complex I (LHI) and/or light-harvesting complex II (LHII), now called B-890 and B-800-850, respectively (Codgell, 1986). LHI occurs at a fixed stoichiometry relative to the RC and is intimately associated with RC, and has one absorption maximum in the near infrared. LHII is present in variable amounts and is located more peripherally around the RC. LHII has two absorption maxima in the near infrared (Cogdell, 1986). Bacteriachlorophyll and carotenoids are non-covalently bound to two types of apoproteins: α- and β-apoproteins. By varying the amount of LHII, bacteria are able to adjust in response to various environmental conditions such as light intensity. Detailed structures of LHII of *Rhodopseudomonas acidophila* (McDermott et al., 1995) and LHI of *Rhodospirilum rubrum* (Karrasch et al., 1995) have been determined by crystallography.

Light-harvesting mechanisms utilized by oxygenic organisms took a different path. For instance, the primitive cyanobacterium ancestor that gave rise to contemporary photosynthesis (Figure 1.2) presumably utilized phycobilisomes as antenna. Phycobilisomes, still present in most contemporary cyanobacteria, are large protein-pigment complexes that contain covalently attached bilins (linear tetrapyrroles). Phycobilisomes are associated with PSII. PSI excitation energy is derived from PSII by a mechanism referred to as "spillover" (Williams and Allen, 1987). At some point, antenna complexes arose that were specifically associated with PSI (Figure 1.2). The most likely ancestor antenna complex was similar to that present in contemporary red algae. These algae still have phycobilisomes that transfer energy to PSII, but, importantly, have chlorophyll-protein complexes specifically associated with PSI (Wolfe et al., 1994). These PSI-associated chlorophyll-protein complexes are thought to be the ancestral

progenitor of light-harvesting proteins present in modern day eukaryotic algae and higher plants (Figure 1.2). Diatom chloroplasts are thought to be descendants of red algae (Figure1.2) through the process of secondary endosymbiosis (Figure 1.2). While the diatom chloroplast is most closely related to red algae, the host eukaryotic "pre-diatom" cell was most closely related to Oomycetes (Gunderson et al. 1987; Medlin et al. 1993).

Purpose of this work

The main objective of my research was to characterize *fcp* genes and proteins, with an emphasis on those that specifically serve photosystem I in the diatom *Cylindrotheca fusiformis*. This diatom has been used as a model to study nuclear microtubles (Cande and MacDonald, 1985), silicon and nitrate uptake (Hildebrand et al., 1998; Hildebrand and Dahlin, 2000), cell wall biogenesis (Kroger 1996), and more recently as a model of photosynthetic processes (Martinson et al., 1998). The approach followed in this thesis involved obtaining the sequences of the *fcp* genes and biochemical characterization of PSI-LHCI complexes, two very different methodologies that required techniques in molecular biology and protein biochemistry.

In the first part of the work, a PSI-LHCI complex was isolated and characterized with respect to its spectroscopic and biochemical properties. Importantly, oxygen-evolving thylakoid membranes (Martinson et al. 1998) were used for this study (see above for comments about the importance of using oxygen-evolving thylakoids). The PSI-LHCI complex contained PsaA/B and PsaC, the later determined by N-terminal AA sequence analysis, suggesting the complex contained all the electron transport components in PSI (Fromme, et al., 2001), and probably most, if not all, of the components required for electron transport from Cyt c_{553} to ferredoxin. The complex also contained nine FCP proteins, as determined using immunostaining. The PSI-LHCI complex lacked detectable PSII core proteins, as determined by immunostaining and silver staining, indicating that all FCP proteins were specifically associated with PSI. Energy transfer from FCP

proteins to PSI was confirmed by 77K fluorescence analysis: a single fluorescence emission peak was observed at 715 nm, likely from the PSI core, but possibly from a FCP with long wavelength pigments. Importantly, there were no fluorescence signals that could be associated with either free pigments (i.e., 676 and 640 nm, for Chl *a* and Chl *c*, respectively) or from dissociated FCP (i.e., 680-683 nm). The complex contained eight pigments. Chl *a* and ß-carotene were present, mostly likely in PSI, while Chl *a*, Chl c_1, Chl c_2, fucoxanthin, neofucoxanthin, diatoxanthin, and diadinoxanthin were present as components of LHCI. The presence of nine FCPs in a PSI-LHCI complex suggests either that the diatom LHCI has more FCPs than is present in plants or green algae (four Lhca proteins) or red algae (5-6) or that some of the FCPs are similar to the mobile LHC of green plants, and are capable of associations with PSI and PSII, depending upon growth conditions. Other interpretations are also discussed.

The second part of this work describes the genomic study of *fcp* genes. Twenty *fcp* genes were cloned and sequenced (see Appendix I). We have cloned and sequenced internal fragments, 5' end fragments, and cDNA fragments. When the deduced amino acid sequences of these genes were compared with the N-terminal sequences of the two most abundant FCPs, the 18kD and the 17kD proteins (Martinson, 1996), there were eight genes encoding the 18kD and six encoding the17kD FCP; six genes encoded other FCPs that have not been identified by N-terminal amino acid sequencing, three of which have very close sequence similarity to a 17.5kD FCP (Martinson, 1996).

References:

Adamska, I. 1997. ELIPS – light-induced stress proteins. *Physiol. Plant.* 100:794-805.

Alt, J., Morris, J., Westhoff, P. & Herrmann, R. G. 1984. Nucleotide sequence of the clustered genes for the 44kD chlorophyll a apoprotein and the "32kd-like" protein of the photosystem II reaction center in the spinach plastid chromosome. *Curr. Genet.* 8:597-606.

Apt, K. E., Clendennen, S. K., Powers, D. A. & Grossman, A. R. 1995. The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*. *Mol. Gen. Genet.* 246:455-464.

Barber, J. & Andersson, B. 1994. Revealing the blueprint of photosynthesis. *Nature* 370:31-34.

Bassi, R., Hoyer-Hansen, G., Barbato, R., Giacometti, G. M. & Simpson, D. J. 1987. Chlorophyll-proteins of the photosystem II antenna system. *J. Biol. Chem.* 262:13333-13341.

Bathke, L., Rhiel, E., Krumbein, W. E. & Marquardt, J. 1999. Biochemical and immunochemical investigations on the light-harvesting system of the Cryptophyte *Rhodomonas sp.*: evidence for a photosystem I specific antenna. *Plant Biol.* 1:516-523.

Baymann, F., Brugna, M., Mülenhoff, U. & Nitschke, W. 2001. Daddy, where did (PS)I come from? *Biochim. Biophys. Acta* 1507:291-310.

Berkaloff, C., Caron, L. & Rousseau, B. 1990. Subunit organization of PSI particles from brown algae and diatoms: polypeptide and pigment analysis. *Photosynth. Res.* 23:181-193.

Bhaya, D. & Grossman, A. R. 1993. Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom *Phaeodactylum triconutum*. *Nucleic Acids Res.* 21:4438-4446.

Blankenship, R. E. 1992. Origin and early evolution of oxygenic photosynthesis. *Photosynth. Res.* 33:91-111.

Blankenship, R. E. & Hartman, H. 1998. The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* 23:94-97.

Boekema, E. J., Jensen, P. E., Schlodder, E., Breemen, J. F. L., Roon, H., Scheller, H. V. & Dekker, J. P. 2001. Green plant photosystem I binds light-harvesting complex I on one side of the complex. *Biochemistry* 40:1029-1036.

Brown, J. S. 1988. Photosynthetic pigment organization in diatoms (Bacillariophyceae). *J. Phycol.* 24:96-192.

Camm, E. L. & Green, B. R. 1983. Fractionation of thylakoid membrane with the nonionic detergent octyl-ß-glucopyranoside: resolution of chlorophyll-protein complex II into two chlorophyll-protein complexes. *Plant Physiol.* 66:428-432.

Cande, W. Z. & McDonald, K. L. 1985. *In vitro* reactivation of anaphase elongation using isolated diatom spindles. *Nature* 316:168-170.

Caron, L. & Brown, J. 1987. Chlorophyll-carotenoid protein complexes from the diatom, *Phaeodactylum triconutum*: spectrophotometric, pigment and polypeptide analyses. *Plant Cell Physiol.* 28:775-785.

Chitnis, P. R. 2001. Photosystem I: function and physiology. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 52:593-626.

Cogdell, R. J. 1986. Light-harvesting complexes in the purple photosynthetic bacteria. *In* Staehelin, L. A. and Arntzen, C. J. [Eds.] *Encyclopedia of Plant Physiology.* Vol. 19. Springer-Verlag, Berlin, Hindelberg, New York, Tokyo, pp. 252-258.

DeMartino, A., Douady, D., Quinet-Szely, M., Rousseau, B., Crepineau, F., Apt, K. & Caron, L. 2000. The light-harvesting antenna of brown algae: highly homologous proteins encoded by a multigene family. *Eur. J. Biochem.* 267:5540-5549.

Dunahay, T. G. & Staehelin, L. A. 1986. Isolation and characterization of PSII core complexes from a brown alga, *Laminaria saccharina. FEBS Lett.* 24:22-26.

Eppard, M. & Rhiel, E. 1998. The genes encoding light-harvesting subunits of *Cyclotella cryptica* (Bacillariophyceae) constitute a complex and heterogeneous family. *Mol. Gen. Genet.* 260:335-345.

Eppard, M., Krumbein, W. E., Haeseler, A. & Rhiel, E. 2000. Characterization of fcp4 and fcp 12, two additional genes encoding light harvesting proteins of *Cyclotella cryptica* (Bacillariophyceae) and phylogenetic analysis of this complex gene family. *Plant Biol.* 2:283-289.

Friedman, A. L. & Alberte, R. S. 1984. A diatom light-harvesting pigment-protein complex. *Plant Physiol.* 76:483-489.

Fromme, P., Jordan, P. & Krauß, N. 2001 . Structure of photosystem I. *Biochim. Biophys. Acta* 1507:5-31.

Funk, C., Schröder, W. P., Green, B. R., Renger, G. & Andersson, B. 1994. The intrinsic 22kDa protein is a chlorophyll binding subunit of photosystem II. *FEBS Lett.* 342:261-266.

Gibbs, S. P. 1970. The comparative ultrastructure of the algal chloroplast. *In* Fredrick, J.F. & Klein, R.M. [Eds] *Annal of the New York Academy of Science.* Vol. 175. Art. 2. New York Academy of Science, New York, pp. 454-473.

Gobets, B. & Grondelle, R. V. 2001. Energy transfer and trapping in photosystem I. *Biochim. Biophys. Acta* 1507:80-99.

Golbeck, J. H. 1993. Shared thematic elements in photochemical reaction centers. *Proc. Natl. Acad. Sci.* USA 90:1642-1646.

Golbeck, J. H. 1994. Photosystem I in cyanobacteria. *In* Bryant, D.A. [Ed] *The Molecular Biology of Cyanobacteria.* Kluwer Academic Publishers, Netherlands, pp. 319-360.

Green, B. R. & Durnford, D. G. 1996. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Physiol. Plant Moil. Biol.* 47:685-714.

Green, B. R. & Kühlbrandt, W. 1995. Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosynth. Res.* 44:139-148.

Green, B. R. & Pichersky, E. 1994. Hypothesis for the evolution of three-helix Chl *a/b* and Chl *a/c* light-harvesting antenna proteins from two-helix and four-helix ancestors. *Photosynth. Res.* 39:149-162.

Green, B. R., Pichersky, E. & Kloppstech, K. 1991. Chlorophyll *a*/*b*-binding proteins: an extended family. *Trends Biochem. Sci.* 16:181-186.

Gunderson, J. H., Elwood, H., Ingold, A., Kindle, K. & Sogin, M. 1987. Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. *Proc. Natl. Acad. Sci. USA* 84:5823-2327.

Harworth, P., Watson, J. L. & Arntzen, C. J. 1983. The detection, isolation, and characterization of a light-harvesting complex which is specifically associated with Photosystem I. *Biochim. Biophys. Acta* 724:151-158.

Hildebrand, M., Dahlin, K. & Volcani, B. E. 1998. Characterization of a silicon transporter gene family in *Cylindrotheca fusiformis* sequences, expression analysis, and identification of homologs in other diatoms. *Mol. Gen. Genet.* 260:480-486.

Hilderbrand, M. & Dahlin, K. 2000. Nitrate transporter genes from the diatom *Cylindrotheca fusiformis* (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle. *J. Phycol.* 36:702-713.

Hill, R. & Bendall, F. 1960. Function of the two cytochrome components in chloroplasts: a working hypothesis. *Nature* 186:136-137.

Hipkins, M. F. & Baker, N. R. 1986. Spectroscopy. *In* Hipkins, M.F. & Baker, N.R.[Eds] *Photosynthesis energy transduction: a practical approach.* IRL Press, Oxford, Washington, pp. 51-102.

Ihalainen, J. A., Gobets, B., Sznee, K., Brazzoli, M., Croce, R., Bassi, R., Grondelle, R. V., Kroppi-Tommola, J. E. I. & Dekker, J. P. 2000. Evidence of two spectroscopically different dimers of light-harvesting complex I from green plants. *Biochemistry* 39:8635-8631.

Ikeuchi, M. 1992. Subunit proteins of photosystem I. *Plant Cell Physiol.* 33:669-676.

Jansson, S. 1994. The light-harvesting chlorophyll *a*/*b*-binding proteins. *Biochim. Biophys. Acta* 1184:1-19.

Jansson, S. 1999. A guide to the *Lhc* genes and their relatives in *Arabidopsis*. *Perspectives* 4:236-240.

Jansson, S., Andersen, B. & Scheller, H. V. 1996. Nearest-neighbor analysis of higher-plant photosystem I holocomplex. *Plant Physiol.* 112: 409-420.

Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W. & Krauß, N. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5Å resolution. *Nature* 411:909-917.

Karrasch, S., Bullough, P. A. & Robin, G. 1995. The 8.5 Å projection map of the light-harvesting complex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. *EMBO J* 14:631-638.

Knoetzel, J. & Simpson, D. 1991. Expression and organization of antenna proteins in the light- and temperature-sensitive barley mutant *chlorina*. *Planta* 185:111-123.

Kroger, N., Bergsdorf, C. & Sumper, M. 1996. Frustulins: domain conservation in a protein family associated with diatom cell walls. *Eur. J. Biochem.* 239:259-264.

Kroth-Pancic, P. G. 1995. Nucleotide sequence of two cDNA encoding fucoxanthin chlorophyll *a/c* proteins in the diatom *Odontella sinensis*. *Plant Mol. Biol.* 27:825-828.

Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. 1994. Atomic model of plant light-harvesting complex by electron crystallography. *Nature* 367:614-621.

LaRoche, J., Henry, D., Wyman, K., Sukenik, A. & Falkowski, P. 1994. Cloning and nucleotide sequence of a cDNA encoding a major fucoxanthin-, chlorophyll *a/c*-containing protein from the chrysophyte *Isochrysis galbana*: implications for evolution of the *cab* gene family. *Plant Mol. Biol.* 25:355-368.

Lancaster, C. R. D., Bibikova, M. V., Sabatino, P., Oesterhelt, D. & Michel, H. 2000. Structural basis of the drastically increased initial electron transfer rate in the reaction center from a *Rhodopseudomonas viridis* mutant described at 2.00-Å resolution. *J. Biol. Chem.* 275:39364-39368.

Li, X-P., Björkman, O., Shih, C., Grossman, A. R., Rosenquist, M., Jansson, S. & Niyogi, K. K. 2000. A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* 403:391-395.

Lichtlé, C., Spilar, A. & Duval, J. C. 1992. Immunogold localization of light-harvesting and photosystem I complexes in the thylakoids of *Fucus serratus* (Phaeophyceae). *Protoplasma* 166:99-106.

Lunde, C., Jensen, P. E., Haldrup, A., Knoetzel, J. & Scheller, H. V. 2000. The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. *Nature* 408:613-615.

Masojidek, J., Droppa, M. and Harváth, G. 1987. Analysis of the polypeptide composition of grana and stroma thylakoids by two-dimensional gel electrophoresis. *Eur. J. Biochem.* 169:283-288.

Martinson, T. A. 1996. *Biochemistry of diatom photosynthetic membranes and pigment-protein complexes*. PhD thesis. University of Alaska Fairbanks, Fairbanks, AK.

Martinson, T. A., Ikeuchi, M. & Plumley, F. G. 1998. Oxygen-evolving diatom thylakoid membranes. *Biochim. Biophys. Acta* 1409:72-86.

McDermott, G., Prince, S. M. & Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J. & Isaacs, N. W. 1995. Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature* 374:517-521.

Medlin, L. K., Williams, D. M. & Sims, P. A. 1993. The evolution of the diatoms (Bacillariophyta) I: origin of the group and assessment of the monophyly of its major divisions. *Eur. J. Phycol.* 28:261-275.

Mullet, J. E., Burke, J. J. & Arntzen, C. J. 1980. Chlorophyll proteins of photosystem I. *Plant Physiol.* 65:814-822.

Mullet, J. E. 1983. The amino acid sequence of the polypeptide segment which regulates membrane adhesion (grana stacking) in chloroplasts. *J. Biol. Chem.* 258:9941-9948.

Nitschke, W. & Rutherford, A. W. 1991. Photosynthetic reaction centers: variations on a common structural theme? *Trends Biochem. Sci.* 16:241-244.

Owens, T. G. 1986. Light-harvesting function in the diatom *Phaeodactylum triconutum* II: distribution of excitation energy between the photosystems. *Plant Physiol.* 80:739-746.

Owens, T. G. & Wold, E. R. 1986. Light-harvesting function in the diatom *Phaeodactylum triconutum*. I: isolation and characterization of pigment-protein complexes. *Plant Physiol.* 80:732-738.

Plumley, F. G. & Schmidt, G. W. 1983. Rocket and crossed immunoelectrophoresis of proteins solubilized with sodium dodecyl sulfate. *Anal. Biochem.* 134:86.

Rhee, K-H., Morris, E. P., Barber, J. & Kühlbrandt, W. 1998. Three-dimensional structure of the plant photosystem II reaction center at 8Å resolution. *Nature* 396:283-286.

Satoh, K. & Bulter, W. L. 1978. Low temperature spectral properties of subchloroplast fractions purified from spinach. *Plant Physiol.* 61:373-379.

Scheller, H. V., Jensen, P. E., Haldrup, A., Lunde, C. & Knoetzel, J. 2001. Role of subunits in eukaryotic photosystem I. *Biochim. Biophys. Acta* 1507:41-60.

Schmid, V. H. R., Cammarata, K. V., Bruns, B. U. & Schmidt, G. W. 1997. *In vitro* reconstitution of the photosystem I light-harvesting complex LHCI-730: heterodimerization is required for antenna pigment organization. *Proc. Natl. Acad. Sci. USA* 94:7667-7672.

Shi, L., Kim, S. J., Marchant, A., Robinson, C. & Schroeder, W. P. 1999. Characterization of the PsbX protein from photosystem II and light regulation of its gene expression in higher plants. *Plant Mol. Biol.* 40:737-744.

Smith, G. J. 1997. Plant gene register RGR 97-099. *Plant Physiol.* 114:1136.

Sprague, A. G. & Varga, A. R. 1986. Topography, composition, and assembly of photosynthetic membranes. *In* Staehelin, L.A. and Arntzen, C.J. [Eds.] *Encyclopedia of Plant Physiology.* Vol. 19. Springer-Verlag, Berlin, Hindelberg, New York, Tokyo, pp. 603-725.

Swiatek, M., Kuras, R., Sokolenko, A., Higgs, D., Olive, J., Cinque, G., Mueller, B., Eichacker, L. A., Stern, D. B., Bassi, R., Herrmann, R. G. & Wollman, F-A. 2001. The chloroplast gene ycf9 encodes a photosystem II (PSII) core subunit, PsbZ, that participates in PSII supramolecular architecture. *Plant Cell* 13:1347-1368.

Ting, C. S. & Owens, T. G. 1993. Photochemical and nonphotochemical fluorescence quenching processes in the diatom *Phaeodactylum triconutum*. *Plant Physiol.* 101:1323–1330.

Williams, W. P. & Allen, J. F. 1987. State1/state2 changes in higher plants and algae. *Photosynth. Res.* 13:19-45.

Wolfe, G. R., Cunningham, F. X., Durnford, D., Green, B. R. & Gantt, E.1994. Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. *Nature* 367:566-568.

Xiong, J., Fischer, W. M., Inoue, K., Nakahara, M. & Bauer, C.E. 2000. Molecular evidence for the early evolution of photosynthesis. *Science* 289:1724-1730.

Zouni, A., Witt, H-T., Kern, J., Fromme, P., Kraußß, N., Saeger, W. & Orth, P. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8Å resolution. *Nature* 409:739-743.

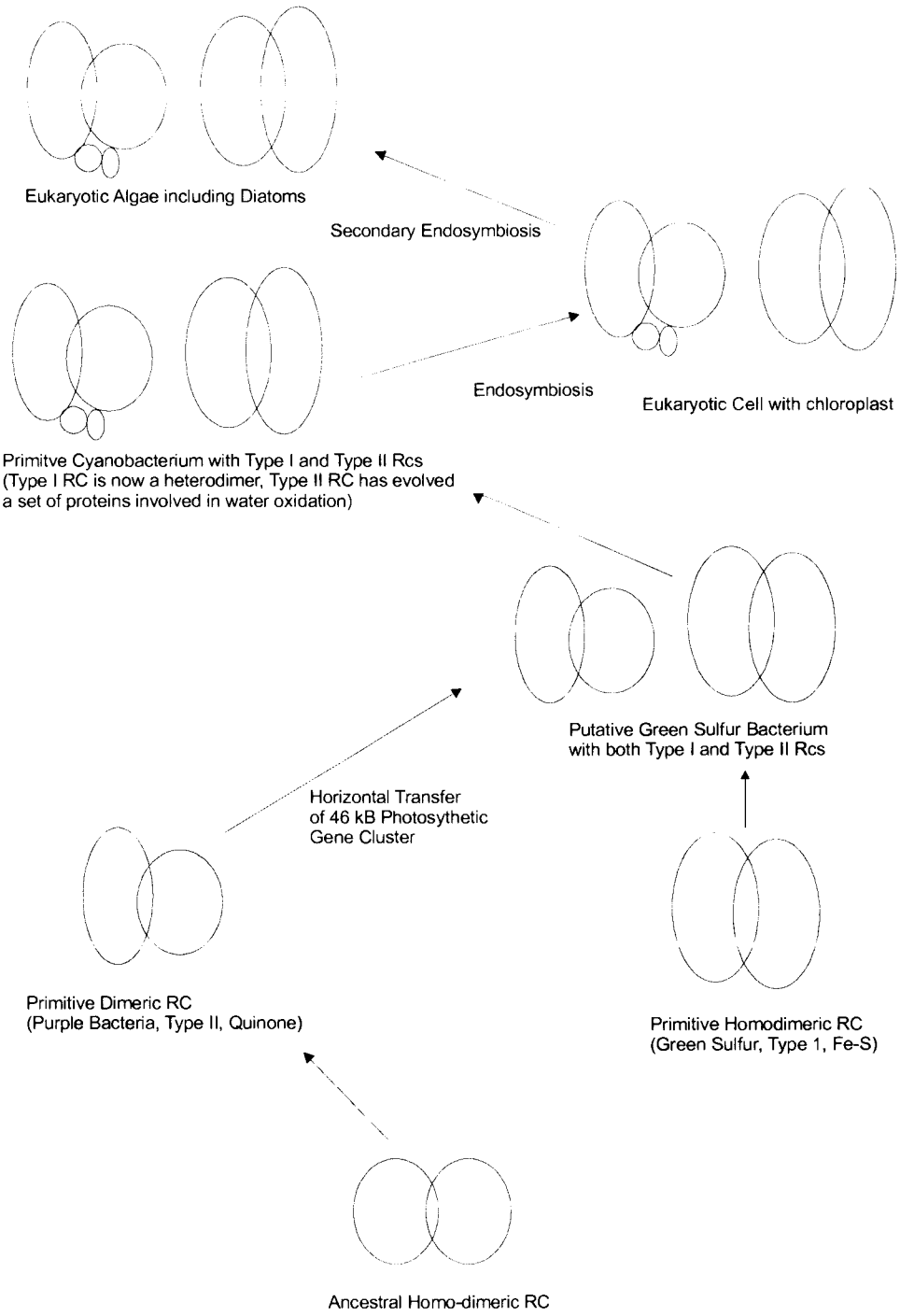

Fig. 1.1 Proposed origin of photosynthetic reaction centers.

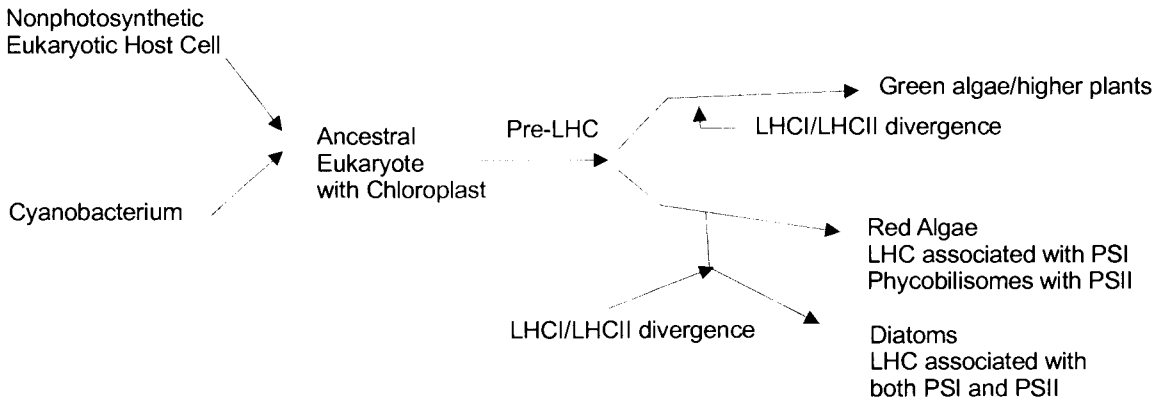

Fig. 1.2. Proposed origin of LHC in algae.

Chapter 2

The PSI-LHCI complex from a Diatom *Cylindrotheca fusiformis*[1]

Hua Zhang[a], F. Gerald Plumley[a,b], Dan Brune[c]

[a] Department of Chemistry and Biochemistry, University of Alaska Fairbanks, Fairbanks, AK 99775
[b] Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, AK99775
[c] Department of Chemistry & Biochemistry, Arizona State University, Box 871604 Tempe, AZ 85287

A PSI-LHCI complex was isolated from oxygen-evolving thylakoids of the diatom *Cylindrotheca fusiformis* following dodecylmaltoside/Lithium docecyl sulfate solubilization and sucrose density gradient centrifugation. The complex was olive brown in color, had a Chl *a/c* ratio of 13 and, relative to thylakoids, was enriched in *ß*-carotene but was depleted in fucoxanthin, Chl c_1 and c_2, and especially diatoxanthin. The CD spectrum of the complex resembled that of the PSI complex of green plants. A single 77K fluorescence emission was observed at 715 nm. The excitation spectra confirmed that both Chl *c* and carotenoids were energetically coupled to Chl *a*. The complex contained PsaA/B and several low molecular weight polypeptides including PsaC as determined by N-terminal amino acid sequencing and nine LHC apoproteins as revealed by immunostaining. PSII core polypeptides were not detected in this preparation by immuno- or silver staining. Taken together, the CD, fluorescence, and protein data indicate that at least nine LHC apoproteins can be specifically associated with PSI in this diatom. Based on comparisons with the phylogenetically related red algal chloroplasts, we predict that six of the antenna proteins belong to LHCI while the remainder are part of the larger LHCII pool that associate with both PSI and PSII in an environmentally sensitive fashion.

[1]Prepared for publication in Journal of Phycology.

Introduction:

Oxygenic photosynthesis is a noncyclic process that utilizes the reducing power generated by the light-driven oxidation of H_2O to produce NADPH. The process involves an array of membrane-bound pigment-protein complexes including two photosystems, photosystem I (PSI) and photosystem II (PSII). Each photosystem is independently activated by light, but electrons flow from PSII to PSI. Photosystem I, consists of a reaction center (RCI), light-harvesting complexes (LHCI), and several accessory subunits that facilitate electron transport. Photosystem I is a highly integrated membrane protein complex that consists of 11 (in cyanobacteria) or 13 (in plants) mostly hydrophobic polypeptides (Ikeuchi, 1992; Fromme, 1996), over 100 antenna pigments, three 4Fe-4S clusters and two phylloquinones. A "native" PSI-LHCI complex was isolated by Mullet et al (1980) from pea (*Pisum sativum)* chloroplasts that contained 110 Chl / P-700. This native complex consisted 10 polypeptides including the reaction center core proteins PsaA/B and had a long wavelength fluorescence emission at 735 nm. Two polypeptides of 24.5 and 22.5kD in this complex were depleted during detergent treatment accompanied with an emission shift to 722nm indicating the presence of LHCI in this complex and the origin of the long wavelength fluorescence. PSI structure and function are well documented in green plants and cyanobacteria including deduced protein sequences of the subunits (Golbeck, 1994; Chitnis et al., 1995) and crystal structure at 2.5Å resolution (Jordan et al., 2001).

Approximately 50% of global ocean production is marine algae, a significant fraction of which is due to Chl *c*-containing algae including chrysophytes, dinoflagellates, brown algae, and diatoms. There are several major differences between the higher plant and the diatom photosynthetic apparatus. Diatom thylakoids are grouped in long stacks of three (Berkaloff et al., 1983) in contrast to the granal and stromal lamellae membranes of green plants. PSI is distributed on the appressed and non-appressed region of the thylakoids but is more concentrated on the two outer non-appressed membranes of the diatom

Phaeodactylum triconutum (Pyszniak and Gibbs, 1992) and the brown alga *Fucus serratus* (Lichtlé, et al., 1992). PSI is localized largely on stromal lamellae and less so on the edges of the grana stacks in green plants (Anderson and Anderson, 1980). The photosynthetic apparatus of both diatoms and green plants contains Chl *a*. However, while the accessory pigments in green plants are Chl *b*, lutein, violaxanthin, and neoxanthin, diatoms have Chl *c*, diatoxanthin, diadinoxanthin, and fucoxanthin as accessory pigments. Xanthophylls serve both light harvesting and structural roles in green plant membranes. It is not known if diatom xanthophylls play a structural role but the molar fraction of fucoxanthin, up to 30% of the total pigment (Gallagher et al., 1984), is much higher in diatoms than in green plants.

Isolation of pigment-protein complexes from diatoms has not met with the same success as the green plant Chl-proteins. The siliceous cell walls are difficult to break without disrupting cellular membranes. Also there is an additional membrane, the chloroplast endoplasmic reticulum (ER), that surrounds and connects to the chloroplast envelope, and diatom thylakoids appear to be more sensitive to detergent treatments than those of green plants. Various P700-containing complexes have been isolated from diatoms (Alberte et al., 1981; Friedman and Alberte, 1984; Owens and Wold, 1986), but they did not contain any LHC and the spectral properties, pigment and polypeptide compositions were not well documented. Two different PSI complexes were isolated from the diatom *Phaeodactylum triconutum* containing 5 or 6 proteins and varying amounts of pigments (Caron and Brown, 1987). Even though small amounts of fucoxanthin and diadinoxanthin as well as Chl *c* were present in the complexes, the direct proof of the existence of LHCI was absent. Both complexes fluoresced at 691-693 nm and 716 nm to various degrees, and the origins of the fluorescence were not clear. Brown (1988) isolated PSI+LHCI fractions from six different diatoms, which all exhibited a broad emission maximum at 716 nm-732 nm along with a minor peak at 676 nm that was characteristic of isolated diatom LHC (Friedman and Alberte, 1984; Owens and Wold, 1986). PSI "native" particles from two brown algae and one diatom (Berkaloff, et al.,

1990) all had a Chl/P700 ratio of about 165 to 375 and exhibited a fluorescence emission maximum at 77K near 720 nm, but were contaminated to varying, but largely unknown, degrees, with PSII and/or bulk LHC.

In this communication, we report the isolation of a PSI-LHCI complex from the diatom *Cylindrotheca fusiformis* using DM and LDS solubilization followed by sucrose density gradient fractionation. Use of O_2-evolving diatom thylakoids (Martinson et al., 1998) as a starting material contributed to the successful isolation of a PSI-LHCI complex that retained its native properties. The CD and fluorescence spectra, along with the lack of PSII core polypeptides revealed by silver staining and immuno-blotting, provided evidence for the existence of a peripheral Chl-protein antenna complex specifically associated with PSI. The antenna contained nine discernable LHC apoproteins. The specific functions of these nine proteins are discussed.

Materials and Methods:

Culture and growth conditions:

Cultures of *Cylindrotheca fusiformis* (Watson's strain [13]) were grown in 4-L Pyrex flasks in artificial seawater (ASW) medium with constant illumination and CO_2-supplemented aeration as described (Martinson et al., 1998).

Formation of protoplast and thylakoid membrane purification using French press:

Formation of protoplasts and purification of thylakoids followed the protocols of Martinson et al. (1998). Briefly, 3.0L of batch culture *Cylindrotheca fusiformis* were harvested by centrifugation and resuspended in 300 ml of an organic-rich protoplast medium (PPM) that contained ASW levels of NaCl but was deficient in Mg^{2+}, Ca^{2+}, Mn^{2+} and silicic acid. Cells were incubated in a 1-L flask in the light with shaking for 24 hours,

during which time cells increased in size approximately 2-3X due to photoheterotrophic nutrition. Cells were unable to complete cell division in PPM due to the absence of Si, which is required for formation of new cell walls. Cells were subsequently harvested by centrifugation and transferred sequentially through a graded series of buffers (Buffers A-D) containing 20 mM phosphate, pH 8.0, protease inhibitors, and varying concentrations of $MgCl_2$ and/or sorbitol (Martinson et al., 1998). Resuspension in Buffer A of low osmotic potential induced the enlarged cells to swell, resulting in loss of the lightly silicified cell walls and the formation of protoplasts. Protoplasts were resuspended in Buffer B with 2 M sorbitol and 20 mM $MgCl_2$ and allowed to equilibrate for 2 hours prior to passage through a chilled French press at very low pressure (2000 psi). Samples were diluted and membrane fractions pelleted by centrifugation in $MgCl_2$-free buffer before flotation through sorbitol step gradients. Thylakoid membrane fractions were collected and resuspended in Buffer C (2M sorbitol). Chlorophylls were quantified spectrophotometrically using the equations of Jeffrey and Humphrey (1975).

Purification of the PSI-LHCI complex:

Thylakoids were solubilized at 0.5mg Chl/ml in 10mM Tris-HCl, pH 8.5, 5 mM ACA, 1 mM BAM with 0.5% (w/v) ß-dodecylmaltoside (DM) for 30 minutes at 0° C in the dark with shaking. Lithium Docecyl Sulphate (LDS) was then added to 0.07% (w/v) and solubilization continued for another 30 minutes. Samples were spun at 16,000 g at 4°C for 10 minutes and the supernatant loaded onto the top of a 5% - 50% sucrose density gradient containing 10 mM Tris-HCl, pH 8.5, 5 mM ACA, 1 mM BAM, 0.01% LDS and spun in a Beckman SW-40 rotor at 230,500 g at 4°C for 17.5 hours. Five pigmented fractions were resolved. The heaviest fraction, olive brown II, sedimented at 30% - 35% sucrose, and was collected and analyzed in this report.

SDS-PAGE:

For denaturing SDS-PAGE, thylakoids were solubilized at 0.5 mg Chl/ml in 0.01 M Tris-HCl, pH 8.5, 0.04 M DTT, 1.25 mM ACA, 0.25 mM BAM, 1.25% LDS, and 7.5% sucrose, heated at 100°C for 1 min, cooled on ice, and 10μg Chl loaded per lane. The sucrose density gradient fractions were concentrated by centrifugation at 95,000 rpm at 4°C for 2 hours in a Beckman TL-100 rotor, and chlorophyll concentrations determined as described above. These fractions were then solubilized in 2% LDS, 125 mM Tris-HCl, pH 8.5, and 125 mM DTT and heated for 1 minute at 100 °C. For dark brown, green I, green II, olive brown I and olive brown II fraction, a total of 16μg, 8μg, 12.8μg, 25μg, and 26 μg Chl were loaded, respectively. Denaturing SDS-PAGE was performed on 10-20% polyacrylamide gels as described (Martinson et al., 1998). After electrophoresis, gels were stained in colloidal Coomassie Brilliant Blue G-250 (Neuhoff et al., 1988). Proteins were also detected by silver staining (Morrissey, 1981). All the washes and incubation were performed at room temperature with shaking. For silver-stained gels, 3μg of Chl was loaded in each lane.

Immunodetection and N-terminal amino acid sequencing:

After denaturing SDS-PAGE, proteins were transferred to nitrocellulose (immuno-blotting) or to PVDF (Millipore, protein sequencing) at 4°C for 1 hour at 20V using a semi-dry transfer apparatus (Bio-Rad). After transfer, the proteins were stained with Ponceau S (15 min; Nitrocellulose) or Amido Black (5 min; PVDF) at room temperature, and destained in ddH_2O (Ponceau S) or 90% methanol, 7% HOAC (Amido black).

For immunodetection, each lane contained 10 μg Chl and blots were incubated with antibodies to CP47 and CP43 from spinach (Martinson et al., 1998), the ATP synthase β-subunits of romane lettuce (Plumley et al., 1989), or the 29 kD LHCP from *Chlamydomas* (Plumley et al., 1993), and developed using alkaline phosphatase-

conjugated goat anti-rabbit IgG in conjunction with 5-bromo-4-chloro-3-indolyl-phosphate and nitroblue tetra-zolium. N-terminal protein sequences were obtained at Arizona State University using a Beckman Coulter Porton 2090E gas-phase protein sequencer (Matsudaira, 1987).

Spectroscopic characterization:

Samples used for absorption and CD analyses were obtained directly from the sucrose density gradient and contained, in different experiments, from 15 –20 μg Chl/ml. Absorption spectra were obtained using an Agilent 8453 diode-array spectrophotometer. CD spectra were measured using a Jasco J-720 spectropolarimeter with 60 μm bandpass. Low temperature (77K) fluorescence spectra were measured in 50% glycerol using an Aminco SLM500-C spectrofluorometer with 4 nm bandpass for both emission and excitation. Data represent the average of at least three scans for both emission and excitation. Emission spectra were corrected to account for low PMT sensitivity in the red region.

Pigment analysis:

Pigments were extracted from the olive brown II fraction with *sec*-butanol as described (Martinson and Plumley, 1995). Pigments were resolved on an Agilent 1100 HPLC using a Waters Symmetry reversed phase C8 column (3.5μm; 4.6 x 150mm) using protocols modified from Zapata et al. (2000).

Results:

Purification of the PSI-LHCI complex:

Sucrose density gradient centrifugation of solubilized thylakoid membranes resulted in the resolution of five membrane fractions: dark brown, green I, green II, olive brown I, and olive brown II, from least to most dense, respectively (Fig. 2.1). The dark brown fractions on top of the gradient (10% - 20% sucrose) fluoresced strongly under UV light and contained mostly light-harvesting proteins and some PSII proteins such as PsbB, PsbC, and PsbO, as well as the ATP synthase β-subunit and PetA (Cyt *f*) (Fig. 2.2a, lane 2). The green I (20% sucrose) and green II (25% sucrose) fractions fluoresced weakly under UV light, and contained different amounts of PsaA/B, ATP synthase β-subunit, PSII core components, light-harvesting proteins as well as other low molecular weight proteins (Fig. 2.2a, lane 3 and lane 4, respectively). The olive brown I fraction (30% sucrose) did not fluoresce under UV light and was enriched in both PSI and PSII polypeptides (Fig. 2.2a, lane 5). The olive brown II fraction (35% sucrose), did not fluoresce under UV light, and was highly enriched in PSI core polypeptides. This fraction did not contain detectable PSII core polypeptides or ATP synthase peptides, as revealed by silver staining and immuno-blotting (Fig. 2.2b lane 2 and Fig 2.2c lane 2). This fraction also contained several low M_r proteins (9-27 kD), which were hypothesized to be PSI subunits and or LHCI apoproteins.

Protein analyses of olive brown II fraction:

Proteins of olive brown II were resolved by denaturing SDS-PAGE and stained with Coomassie blue G250 (Fig. 2.2a), silver staining (Fig. 2.2b), and immunoblotting (Fig. 2.2c). An abundant protein with a M_r of 65kD cross-reacted with the antibody to the PSI reaction center polypeptide (PsaA/B) (data not shown). No ATP synthase subunits were detected by silver staining (Fig 2.2b, lane 2) or immunoblotting (Fig 2.2c, lane 2) in the

olive brown II fraction. The olive-brown II fraction was depleted in PSII components such as PsbB, PsbC, PsbO, D1, and D2, as revealed by silver staining (Fig 2.2b, lane 2). No photosystem II core proteins were present when olive brown II was hybridized to antibodies against CP 47 (PsbB), and CP43 (PsbC) of spinach (data not shown). When blots were challenged with an antibody raised against the 29-kD LHCP of *Chlamydomonas reinhardtii*, nine bands cross-reacted (Fig 2.2c, lane 2). These bands have apparent molecular weights of 27, 22, 21, 20, 18, 17.5, 16, 14, and 10kD. The N-terminal sequence of a 9 kD protein was MSHTVKIYDT. BLAST analysis revealed this sequence to be homologous to PsaC proteins of cyanobacteria, higher plants, and diatoms.

Pigment analyses and spectral characterization of olive brown II fraction:

The room temperature absorption spectrum of olive brown II, the putative PSI-LHCI complex, exhibited two major peaks at 439 nm and 680 nm that were characteristic of Chl *a*, and shoulders at 469 nm and 630 nm characteristic of Chl *c* (Fig. 2.3). The shoulder at 498 nm and the broad region of absorption between 480 nm and 580 nm is likely due to fucoxanthin and other carotenoids such as diatoxanthin and diadinoxanthin (Anderson and Barrett, 1979) and β-carotene (Owens and Wold, 1986). The pigment compositions of the olive brown II and thylakoids were compared using HPLC (Fig 2.4) with chromatograms normalized to Chl *a*. Thylakoids and the olive brown II fraction both contained Chl *a*, Chl c_1 and c_2, fucoxanthin, neofucoxanthin, diadinoxanthin, diatoxanthin, and ß-carotene. Compared to thylakoids, olive brown II contained reduced levels of all accessory pigments, but was enriched in ß-carotene, as expected for a fraction that was enriched in PSI but depleted in bulk LHC.

Low temperature fluorescence spectroscopy was used to investigate the organization of Chls and carotenoids in the olive brown II fraction. Excitation at 440, 470, and 490 nm was used to preferentially excite Chl *a*, Chl *c*, and fucoxanthin, respectively. Regardless

of the excitation wavelength, the fluorescence emission spectra exhibited a single emission peak centered at 715 nm (Fig. 2.5a) indicating that accessory pigments were effectively coupled to Chl *a*. The fluorescence excitation spectrum for the 715 nm emission (Fig. 2.5b) revealed excitation bands at 439, 469, and 490 nm, and one broad region at 570-590 nm. Emissions driven by excitation at 440, 470 and 490 nm are characteristic of Chl *a*, Chl *c*, and fucoxanthin absorption bands, respectively, and demonstrate a high degree of energy transfer from fucoxanthin and Chl *c* to Chl *a*. The broad region between 570 and 590 nm was probably due primarily to fucoxanthin.

Pigment-pigment and/or pigment-protein interactions in olive brown II were investigated using circular dichroism spectroscopy. Overall, the CD spectrum of the olive brown II complex (Fig. 2.6) resembled the PSI complex of green plants (Nabedryk et al., 1984). We observed a 678 (-)/660 (+) excitonic signal that is characteristic of *Cylindrotheca* CPI (Martinson, 1996), and a 464(-)/444(+)/434(-) feature similar to the one reported in the light-harvesting complexes of another chlorophyll *c*-containing alga (Büchel and Garab, 1997). A 464(-) nm peak was also observed in green plant PSI particles (Ikegami and Itoh, 1986), making the identification of this feature less certain.

Discussion:

A PSI-LHCI complex has been isolated and characterized from the diatom *C. fusiformis* using O_2-evolving thylakoids (Martinson et al., 1998) as the starting material. The complex contained PsaA/B, PsaC, and nine FCPs that were energetically coupled to PSI. No PSII core components were detected by immuno- or silver staining, thus excluding the possibility that the FCP proteins were transferring energy to PSII and then to PSI via spillover. Based on the LHCI composition of green plants and red algae, we speculate that six of the identified FCPs are components of LHCI, while the remaining FCPs are part of the larger, and perhaps mobile, antenna complex of this diatom.

Fluorescence properties of the PSI-LHCI complex

Low temperature fluorescence spectroscopy can be used to measure the distribution of absorbed light energy between LHCs and reaction centers. The PSI-LHCI fraction of *Cylindrotheca* had a single 77K fluorescence emission maximum at 715 nm, very close to the wavelength reported for PSI from whole cells (714 nm) and thylakoids (717 nm) (Martinson, et al., 1998). PSI-enriched fractions have been isolated from several chromophytes with very different protein compositions and fluorescence properties (Berkaloff et al., 1990; Caron and Brown, 1987; Brown, 1988). Two PSI-containing fractions from *Phaeodactylum triconutum* fluoresced at 691 nm and 716 nm (Caron and Brown, 1987); the fraction that was more enriched in P700 fluoresced mainly at 691 nm. In contrast, 'native' PSI fractions from *Phaeodactylum triconutum* and two brown algae all had a single fluorescence emission maximum at about 720nm (Berkaloff et al., 1990). There are several reasons for the dramatically different fluorescence properties of these PSI fractions. Detergents must be used to solubilize thylakoids and each detergent potentially affects fluorescence properties in a different fashion. Moreover, detergent pigment ratios are also a cause of potential problems. For instance, detergents have been shown to induce shifts to shorter wavelength in the fluorescence spectrum of 'native' PSI fractions (Berkaloff et al., 1990) and affect the fluorescence emission of other pigment-proteins complexes in Chl *c*-containing algae (Katoh et al., 1989; Büchel and Wilhelm, 1993). Another reason that PSI fluorescence properties have been difficult to understand results from the frequent contamination of PSI fractions by PSII and/or LHCII (Caron and Brown, 1987; Berkaloff et al., 1990). For instance, a PSI-enriched fraction of *P.triconutum* that contained PSII contamination, had an emission maximum at 691nm, which was very close to the fluorescence maximum of the PSII-enriched fraction (693nm). Other contaminants in PSI fractions include LHCII, which in theory, may be associated with PSII or PSI. A final complicating issue is that PSI-LHCI fractions from green plants generally have a long wavelength emission at 730-735 nm that arises from pigments within LHCI, rather than PSI (Jansson, 1994). Long wavelength fluorescence

emissions from LHCI are thought to be restricted to higher plants, but it is not certain if the 715 nm signal from diatom is from PSI or LHCI. The PSI-LHCI fraction of *Cylindrotheca* had a single fluorescence maximum at 715 nm (Fig 2.4a), which was very close to the fluorescence maximum of several diatoms (Brown, 1987) and Chl *c*-containing algae (Büchel and Wilhelm, 1993; Berkaloff et al., 1990), indicating energy transfer from LHCI proteins to PSI. Importantly, there were no fluorescence signals that could originate from either free pigments (i.e., 676 and 640 nm, for Chl *a* and Chl *c*, respectively) or from dissociated FCP (i.e., at 680-683 nm). More importantly, the diatom PSI fraction was depleted of PSII components (Fig. 2.2a, Fig. 2.2b). LHCII, if present, was energetically coupled to PSI.

Light-harvesting proteins of the PSI-LHCI complex

When the PSI-LHCI fraction was challenged with antibody raised against the *Chlamydomonas* 29 kD LHCP, nine protein bands ranging from 27 kD to 10 kD cross-reacted (Fig 2.2C). There are only four to six LHCPs that specifically serve PSI in green plants and five or six in red algae (Jansson, 1999; Marquardt et al., 2000; Marquardt et al., 2001; Tan et al., 1997a; Tan et al., 1997b). Ten LHCPs were identified in a Cryptophyte, *Rhodomona sp.*, with at least three of them associated specifically with PSI (Bathke et al., 1999). Information about LHCPs of PSI is limited, mostly because they are difficult to isolate abundantly in an intact state (Schmid et al., 1997). It is widely accepted that LHCI consists of four polypeptides (Lhca 1-4) in green plants that have protein masses in the range of 20-24 kD. The respective genes have also been identified, cloned, and sequenced (Jansson, 1994). It was suggested that LHCI forms two kinds of heterodimers in solution, LHCI-730 and LHCI-680. Detailed spectroscopic characterization of each component is available (Ihalaninen et al., 2000). LHCI binds PSI on only one side (Boekema et al., 2000). No evidence is available from other photosynthetic organisms with LHCPs, so it is difficult to know if the diatom LHCI has a similar organization with PSI.

While the precise LHCI protein composition of diatoms is not known, we argue that six of the nine FCPs present in the diatom PSI-LHCI is specific for PSI while the others may serve both PSI and PSII. Since the LHCPs associated with PSI appear to have diverged relatively early in evolutionary history (Wolfe et al., 1994), their stoichiometry and interaction with PSI may differ significantly between species (Scheller et al., 2001). In other words, there could be more than six FCPs that serve PSI in diatoms, but it seems reasonable that at least some of the FCPs (Fig 2.2C) are members of LHCII. There are, however, other possibilities, and these will be discussed.

One explanation for the presence of nine detectable FCPs (Fig. 2.2C) is that LHCPs from green plants are known to undergo post-translational modification (Jansson, 1994) that could result in differential migration during SDS-PAGE (Berganotino et al., 1995). The green plant/alga LHCPII apoproteins are post-translationally phosphorylated, a process that is regulated by chloroplast redox levels, and affects the three-dimensional structure (Nilsson et al. 1997). Other post-translational modifications in green plants include amino acid methylation, acetylation, and fatty acid acylation. More specific to LHCPI, 18 LHCPIs were separated by high-resolution two-dimensional gel electrophoresis from an isolated PSI of *Chlamydomonas reinhardtii* (Hippler et al., 2001), several of which proved to be products arising from post-translational modifications.

We don't know whether FCPs of diatoms are subject to the same post-translational modifications as green plant LHCPs, however, N-terminal sequencing revealed a hydroxy-proline residue in both the 18 kD and the 17 kD FCP of *Cylindrotheca fusiformis* (Martinson, 1996; Martinson et al., in prep). Preliminary *in vivo* phosphorylation studies of *Cylindrotheca* indicate little, if any phosphorylation of the abundant 18 and 17 kD FCPs. Several proteins in the size range of the proteins reported here (10-27 kD; Fig. 2.2C) were phosphorylated, though it is not known if these proteins were associated with PSI.

Functions of the FCPs in the PSI-LHCI complex

Light-harvesting complexes have been isolated and characterized from the diatom *Phaeodactylum triconutum* (Friedman and Alberte, 1984; Fawley and Grossman, 1986; Caron and Brown, 1987) but the amino acid sequences of the FCPs were not determined. On the other hand, the DNA sequences of LHCP from several Chl *c*-containing algae have been determined (Grossman et al., 1990; Hiller et al., 1995; Paassaquet and Lichtle 1995; Apt et al., 1995; Durnford and Green, 1995; Kroth-Pancic, 1995), but it is not clear if the cognate apoproteins are associated with PSI or PSII. Previous results (Plumley et al., 1993) indicated that there were at least four FCPs (18, 17.5, 17, and 16.5kD) in *Cylindrotheca*. More detailed analysis, including use of thylakoids prepared by different procedures, showed that two additional polypeptides, 27 and 20kD, were also present (Martinson et al., 1998). Both of these studies relied upon the use of an antibody raised against a *Chlamydomonas* LHCP, the same antibody that was used in this work. Possible reasons for the differences between our results (nine bands) and previous results include the use of a different batch of antibody that recognized additional epitopes and/or the increased purity of the PSI-LHCI fraction.

Twenty *fcp* genes including ten full-length cDNA have been cloned and sequenced from *C. fusiformis* (see Appendix I). The precise function of each gene is not known. The deduced amino acid sequences of eight of the genes match the N-terminal sequences of an 18 kD FCP and six match a 17 kD FCP (Martinson, 1996). The deduced amino acid sequences of three genes are very close to the N-terminal sequence of a less abundant 17.5 kD FCP. Together, the 18 and 17 kD FCPs are the most abundant FCPs in *C.fusiformis*, comprising 60-80% of the total FCP biomass (Martinson et al., in prep). Although we lack conclusive evidence, it is likely that the 18 and 17 kD FCPs serve both PSI and PSII; otherwise there would be an imbalance in excitation energy transfer to the reaction centers. Among the nine FCPs identified in our PSI-LHCI (Fig. 2.2c), we note that an 18 kD protein band cross-reacted weakly while the strongest band had an apparent

molecular weight of 17.5 kD. There was no obvious cross-reaction with a 17 kD protein. Attempts to obtain the N-terminal sequences of the nine FCPs from our PSI-LHCI fraction yielded data that were too inconclusive to definitely determine if any of the FCPs matched our gene sequences (Appendix 1) or previously published N-terminal sequences (Martinson 1996). However, obtaining the N-terminal sequences of the FCPs associated with PSI should contribute to current ongoing efforts towards establishing a preliminary picture of the structure and function of the light-harvesting protein family from a diatom, as current information is available only for green plants, one class of green algae, and a red alga.

PsaC in the PSI-LHCI complex

The PSI-LHCI fraction of *Cylindrotheca* also contained numerous low molecular weight proteins, including PsaC, a 9 kD hydrophilic protein. PsaC is located on the stromal side of the thylakoids and contains two 4Fe-4S clusters that serve as the final electron acceptor in the electron PSI transport chain (Fromme, et al., 2001). Electron transfer within PSI starts with an excited chlorophyll (Chl^*) that flows to P700→ A_0 → A_1 → F_x → F_A → F_B, where F_A and F_B are the two 4Fe-4S clusters on PsaC. All of the cofactors required for charge separation and charge stabilization are located on the three polypeptides encoded by the *psa*A, *psa*B, and *psa*C genes (Golbeck, 1992). Removal of PsaC results in the loss of the F_A/F_B clusters from the PSI reaction center, with F_x becoming the terminal electron acceptor (Golbeck, 1992). PsaC transfers electrons to ferredoxin, a mobile carrier capable of reducing several different substrates, but most electrons flow to NADP and then to CO_2. In addition to its role in electron transfer, PsaC has been reported as an essential component for assembly of the PSI complex in *Chlamydomonas reinhardtii* (Takahashi et al., 1991) and the level of PsaC has been used as a proxy that accurately reflects the number of PSI reaction centers (Jansson et al., 1997). The green plant PsaC can be easily removed from the PSI complex with chaotropic agents, such as urea, or LDS. However, the use of LDS did not remove PsaC

from the diatom PSI-LHCI complex (Fig. 2.2a). We are uncertain why diatom PsaC was not removed by LDS, but note that the LDS concentration used here was very low (0.07%). PsaC was much more easily dissociated from cyanobacteria PSI than from plant PSI and the difference appeared to be related to the different structure of PsaD rather than to PsaC itself or to the PsaA/B heterodimer (Scheller et al., 2001). PsaD has been identified by N-terminal sequencing in *Cylindrotheca* (Martinson, 1996) and it has a molecular weight of 14 kD that is different from that of green plant (17.6 kD) or cyanobacteria (15.6 kD). The failure of LDS treatment to remove PsaC could also reflect a stronger interaction between PsaC and the PsaA/B complex in diatoms, and/or be a result of the different thylakoid ultrastructure present in diatoms as compared to green plants. Regardless of the molecular mechanism that allows PsaC recovery as a component of the diatom PSI-LHCI complex, the presence of PsaC confirms the high degree of overall structural integrity of the complex, and lends additional support to our argument that a functional PSI-LHCI complex has been isolated.

Summary:

A PSI-LHCI complex was isolated and characterized with respect to its spectroscopic and biochemical properties using oxygen-evolving thylakoid membranes (Martinson et al. 1998) as the starting material. The PSI-LHCI complex contained PsaA/B and PsaC, the latter determined by N-terminal amino acid sequence analysis, as well as nine FCP proteins detected by immunostaining. The PSI-LHCI complex lacked detectable PSII core proteins, as determined by immunostaining and silver staining, indicating that all FCPs were specifically associated with PSI. Energy transfer from FCP to PSI was confirmed by a single 77K fluorescence emission maximum at 715 nm, and the lack of room temperature fluorescence under UV light. Importantly, there were no fluorescence signals that could be associated with either free pigments (i.e., 676 and 640 nm, for Chl *a* and Chl *c*, respectively) or from dissociated FCP (i.e., at 680-683 nm). The presence of nine FCPs in a PSI-LHCI complex suggests either that the diatom LHCI has more FCPs

than are present in plants or green algae (four Lhca proteins) or red algae (5-6), that the LHCPIs of diatoms undergo extensive post-translational modifications (e.g., phosphorylation, acetylation), or that some of the FCPs are similar to the mobile LHC of green plants, and are capable of associations with PSI and PSII, depending upon growth conditions.

References:

Anderson, B. & Anderson, J. M. 1980. Laternal heterogeneity in the distribution of the chlorophyll-protein complexes of the thylakoid membranes of spinach chloroplasts. *Biochim. Biophys. Acta* 593:427-440.

Anderson, J. M. & Barrett, J. 1986. Photosynthetic membranes and light harvesting systems. *In* Staehelin, L. H. and Arntzen, C. J. [Eds] *Photosynthesis III.* Springer, Berlin, pp. 269-285.

Alberte, R., Friedman, A., Gustafson, D., Rudnick, M. & Lyman, H. 1981. Light-harvesting systems of brown algae and diatoms. Isolation and characterization of chlorophyll *a/c* and chlorophyll a/fucoxanthin pigment-protein complexes. *Biochim. Biophys. Acta* 635:304-316.

Apt, K. E., Clendennen, S, K., Powers, D. A. & Grossman, A. R. 1995. The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera. Mol. Gen. Genet.* 246:455-464.

Bathke, L., Rhiel, E., Krumbein, W. E. & Marquardt, J. 1999. Biochemical and immunochemical investigations on the light-harvesting system of the Cryptophyte *Rhodomonas sp.*: evidence for a photosystem I specific antenna. *Plant Biol.* 1:516-523.

Berkaloff, C., Duval, J. C., Hauswirth, N. & Rousseau, B. 1983. Freeze fracture study of thylakoids of *Fucus Serratus. J. Phycol.* 19:96-100.

Berkaloff, C., Caron, L. & Rousseau, B. 1990. Subunit organization of PSI particles from brown algae and diatoms: polypeptide and pigment analysis. *Photosynth. Res.* 23:181-193.

Bergatino, E., Dainese, P., Cerovic, Z., Sechi, A. & Bassi, R. 1995. A post-translational modification of the photosystem II subunit CP29 protects maize from cold stress. *J. Biol. Chem.* 270:8474-8481.

Boekema, E. J., Jensen, P. E., Schlodder, E., Breemen, J. F. L., Roon, H., Scheller, H. V. & Dekker J. P. 2001. Green plant photosystem I binds light-harvesting complex I on one side of the complex. *Biochemistry* 40:1029-1036.

Brown, J. S. 1988. Photosynthetic pigment organization in diatoms (Bacillariophyceae). *J. Phycol.* 24:96-192.

Büchel, C. & Wilhelm, C. 1993. Isolation and characterization of a photosystem I-associated antenna (LHCI) and a photosystem I-core complex from the chlorophyll *c*-containing alga *Pleurochloris meiringensis* (Xanthophyceae). *J. Photochem. Photobio. B Biol.* 20:87-93.

Caron, L. & Brown, J. 1987. Chlorophyll-carotenoid protein complexes from the diatom, *Phaeodactylum triconutum*: spectrophotometric, pigment and polypeptide analyses. *Plant Cell Physiol.* 28:775-785.

Chitnis, P. R., Xu, Q., Chitnis, V. P. & Nechustai, R. 1995. Function and organization of photosystem I polypeptides. *Photosynth. Res.* 44:23-40.

Durnford, D. G. & Green, B. R. 1995. Characterization of a cDNA encoding a fucoxanthin-chlorophyll protein from the chromophyte alga *Heterosigma carterae*. *In*: Mathis, P. [Ed.] *Photosynthesis: from light to biospere.* Vol I. Kluwer Academic Publisher, Dordrecht, the Netherlands, pp 963-966.

Fawley, M. W. & Grossman, A. R. 1986. Polypeptides of a light-harvesting complex of the diatom *Phaeodactylum triconutum* are synthesized in the cytoplasm of the cell as precursors. *Plant Physiol.* 81:149-155.

Friedman, A. L. & Alberte, R. S. 1984. A diatom light-harvesting pigment-protein complex. *Plant Physiol.* 76:483-489.

Fromme, P. 1996. Structure and function of photosystem I. *Curr. Opin. Struct. Biol.* 6: 473-484.

Fromme, P., Jordan, P. & Krauß, N. 2001. Structure of photosystem I. *Biochim. Biophys. Acta* 1507: 5-31.

Gallagher, J. C., Wood, A. M. & Alberte, R. S. 1984. Ecotypic differentiation in the marine diatom *Skeletonema costatum*: influence of light intensity on the photosynthetic apparatus. *Mar. Biol.* 82:121-134

Golbeck, J. H. 1992. Structure and function of photosystem I. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43:293-324.

Golbeck, J. H. 1994. Photosystem I in cyanobacteria. *In* Bryant, D.A. [Ed.] *The Molecular Biology of Cyanobacteria.* Kluwer Academic Publishers, Netherlands, pp. 319-360.

Grossman, A. R., Manodori, A. & Snyder, D. 1990. Light-harvesting proteins of diatoms: the relationship to the chlorophyll *a/b* binding-proteins of higher plants and their mode of transport into plastids. *Mol. Gen. Genet.* 224:91-100.

Hiller, R. G., Wrench, P. M. & Sharples, F. P. 1995. The light-harvesting chlorophyll *a-c*-binding protein of dinoflagellates: a putative polyprotein. *FEBS Lett* 363:175-178.

Hippler, M., Klein, J. Fink, A., Allinger, T. & Hoerth, P. 2001. Towards functional proteomics of membrane protein complexes: analysis of thylakoid membranes from *Chlamydomonas reinhardtii. Plant J.* 28:595-606.

Ihalainen, J. A., Gobets, B., Sznee, K., Brazzoli, M., Croce, R., Bassi, R., Grondelle, R. V., Kroppi-Tommola, J. E. I. & Dekker, J. P. 2000. Evidence of two spectroscopically different dimers of light-harvesting complex I from green plants. *Biochemistry* 39:8635-8631.

Ikegami, I. & Itoh, S. 1986. Chlorophyll organization in P700-enriched particles isolated from spinach chloroplasts: CD and absorption spectroscopy. *Biochim. Biophys. Acta* 851:75-85

Ikeuchi, M. 1992. Subunit proteins of photosystem I. *Plant Cell Physiol.* 33:669-676.

Jansson, S. 1994. The light-harvesting chlorophyll *a/b*-binding proteins. *Biochim. Biophys. Acta* 1184:1-19.

Jansson, S. 1999. A guide to the *Lhc* genes and their relatives in *Aradidopsis. Trends Plant Sci.* 4:236-240.

Jansson, S., Stefásson, H., Nyström, Gustafsson, P. & Albertsson, P. 1997. Antenna protein composition of PSI and PSII in thylakoid subdomains. *Biochim. Biophys. Acta* 1320:297-309.

Jeffrey, S. W. & Humphrey, G. F. 1975. New spectrophotometric equations for determining chlorophylls *a*, *b*, *c*1, and *c*2 in higher plants, algae, and natural phytoplankton. *Biochim. Physiol. Pflanz.* (BPP) 167:191-194.

Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W. & Krauß, N. 2001. Three – dimensional structure of cyanobacteria photosystem I at 2.5Å resolution. *Nature* 411:909-917.

Kroth-Pancic, P. G. 1995. Nucleotide sequence of two cDNA encoding fucoxanthin chlorophyll *a/c* proteins in the diatom *Odontella sinensis*. *Plant Mol. Biol.* 27:825-828.

Lichtlé, C., Spilar, A. & Duval, J. C. 1992. Immunogold localization of light-harvesting and photosystem I complexes in the thylakoids of *Fucus serratus* (Phaeophyceae). *Protoplasma* . 166:99-106.

Marquardt, J., Wans, S., Rhiel, E., Randolf, A. & Krumbein, W. E. 2000. Intron-exon structure and gene copy number of a gene encoding for a membrane-intrinsic light-harvesting polypeptide of the red alga *Galdieria sulphuraria. Gene* 255:257-265.

Marquardt, J., Lutz, B., Wans, S., Rhiel, E. & Krumbein, W. E. 2001. The gene family coding for the light-harvesting polypeptides of photosystem I of the red alga *Galdieria sulphuraria*. *Photosynth. Res.* 68:121-130.

Martinson, T. A. 1996. *Biochemistry of diatom photosynthetic membranes and pigment-protein complexes*. PhD t hesis. University of Alaska Fairbanks, Fairbanks, AK.

Martinson, T. A., Ikeuchi, M. & Plumley, F. G. 1998. Oxygen-evolving diatom thylakoid membranes. *Biochim. Biophys. Acta* 1409:72-86.

Martinson, T. A. & Plumley, F. G. 1995. One-step extraction and concentration of pigments and acyl lipids by *sec*-butanol from *in vitro* and *in vivo* samples. *Anal. Biochem.* 228:123-130.

Martinson, T. A., Zhang, H., Brune, D., Ikeuchi, M. & Plumley, F. G. in prep. Oligomeric organization of the two major fucoxanthin chlorophyll *a/c* protein complexes in the diatom *Cylindrotheca fusiformis*: is proline hydroxylation involved?

Matsudaria, P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* 262:10035-10038.

Morrissey, J. H. 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. *Anal. Biochem.* 117:307-310.

Mullet, J. E., Burke, J. J. & Arntzen , C. J. 1980. Chlorophyll proteins of photosystem I. *Plant Physiol.* 65:814-822.

Nabedryk, E., Biaudet, P., Darr, S., Arntzen, C. J. & Breton, J. 1984. Conformation and orientation of chlorophyll-proteins in photosystem I by circular dichroism and polarized infrared spectroscopies. *Biochim. Biophys. Acta* 767:640-647.

Neuhoff, V., Arnold, N., Taube, D. & Ehrhardt, W. 1988. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using coomassie brilliant G-250 and R-250. *Electrophoresis* 9:255-262.

Nilsson, A., Stys, D., Drakenberg, T., Spangfort, M. D., Forsen, S. & Allen, J. F. 1997. Phosphorylation controls the three-dimensional structure of plant light harvesting complex II. *J. Biol. Chem.* 272:18350-18357.

Owens, T. & Wold, E. 1986. Light-harvesting function in the diatom *Phaeodactylum triconutum*. I: isolation and characterization of pigment-protein complexes. *Plant Physiol.* 80:732-738.

Passaquet, C. & Lichtlé, C. 1995. Molecular study of a light-harvesting apoprotein of *Giraudyopsis stellifer* (Chrysophyceae). *Plant Mol. Biol.* 29:135-148.

Plumley, F. G., Douglas, S. E., Branagan-Switzer, A. & Schmidt, G. W. 1989. Photosynthesis. *In*: Briggs, W. [Ed.] *Plant Biology*. Alan R. Liss, New York, pp 311-329.

Plumley, F. G., Martinson, T. A., Herrin, D. L., Ikeuchi, M. & Schmidt, G. W. 1993. Structual relationships of the photosystem I and photosystem II chlorophyll *a/b*, and *a/c* light-harvesting apoproteins of plants and algae. *Photochem. Photobiol.* 57:143-151.

Pyszniak, A. M & Gibbs, S. P. 1992. Immunocytochemical localization of photosystem I and the fucoxanthin-chlorophyll *a/c* light-harvesting complex in the diatom *Phaeodactylum triconutum*. *Protoplasma* 166:208-217.

Scheller, H. V., Jensen, P. E., Haldrup, A., Lunde, C. & Knoetzel, J. 2001. Role of subunits in eukaryotic photosystem I. *Biochim. Biophys. Acta* 1507:41-60.

Schmid, V. H. R., Cammarata, K. V., Bruns, B. U. & Schmidt, G. W. 1997. *In vitro* reconstitution of the photosystem I light-harvesting complex LHCI-730: heterodimerization is required for antenna pigment organization. *Proc. Natl. Acad. Sci. USA* 94:7667-7672.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S-Y., Franzén, L. G. & Rochaix, J. D. 1991. Directed chloroplast transformation in *Chlamydomonas reinhardtii*: insertion inactivation of the *psa*C gene encoding the iron sulfur protein destabilizes photosystem I. *EMBO. J.* 10:2033-2040.

Tan, S., Cunningham, F. X., Gantt, E. 1997a. LhcaR1 of the red alga *Porphyridium cruentum* encodes a polypeptide of the LHCI complex with seven potential chlorophyll *a*-binding residues that are conserved in most LHCs. *Plant Mol. Biol.* 33:157-167.

Tan, S., Ducret, A., Aebersold, R. & Gantt, E. 1997b. Red algal LHC I genes have similarities with both Chl *a*/*b*- and *a*/*c*-binding proteins: A 21 kDa polypeptide encoded by LhcaR2 is one of the six LHC I polypeptides. *Photosynth. Res.* 53:129-140.

Wolfe, G. R., Cunningham, F. X., Durnford, D., Green, B. R. & Gantt, E. 1994. Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. *Nature* 367:566-568.

Zapata, M., Rodríguez, F. & Garrido, J. L. 2000. Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C_8 Column and pyridine-containing mobile phases. *Mar. Ecol. Prog. Ser.* 195:29-45.

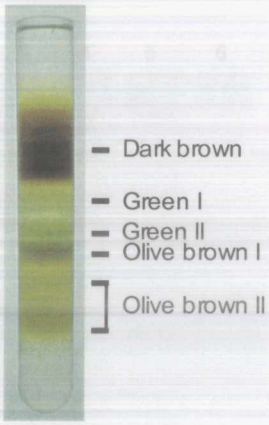

Fig 2.1. Sucrose density gradient of DM/LDS-solubilized thylakoids of *Cylindrotheca fusiformis*.

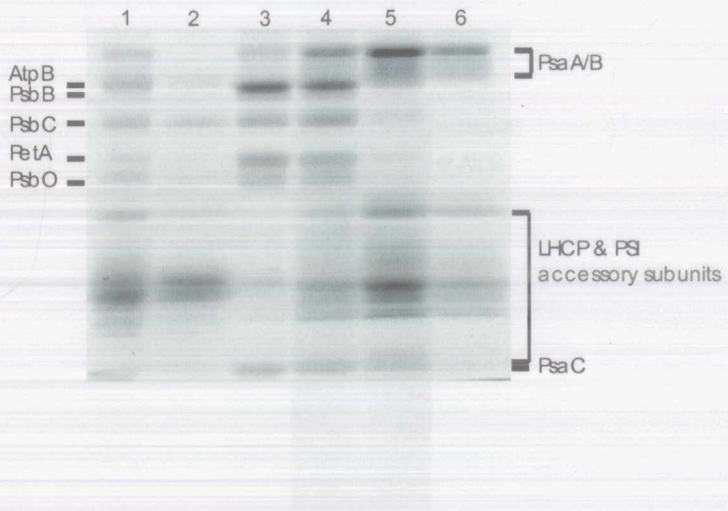

Fig 2.2 a: Coomassie blue G250-stained SDS-PAGE of thylakoids and sucrose density gradient fractions obtained from DM and LDS-solubilized thylakoids. Lane 1: thylakoids; Lane 2: dark brown; Lane 3: green I; Lane 4: green II; Lane 5: olive brown I; Lane 6: olive brown II.

Fig. 2.2b. Silver-stained SDS-PAGE of thylakoids and olive brown II. Lane 1: thylakoids; lane 2: olive brown II. 3μg of Chl was loaded on each lane.

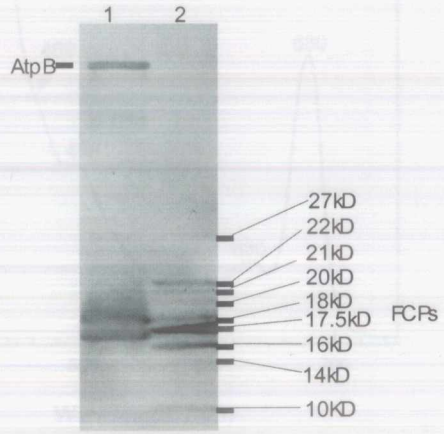

Fig. 2.2c. Immunoblotting of thylakoids (lane 1) and olive brown II (lane 2) using antibodies to ATP synthase ß-subunits from romain lettuce and 29kD LHCP from *Chlamydomonas*. 10 μg of Chl was loaded on each lane.

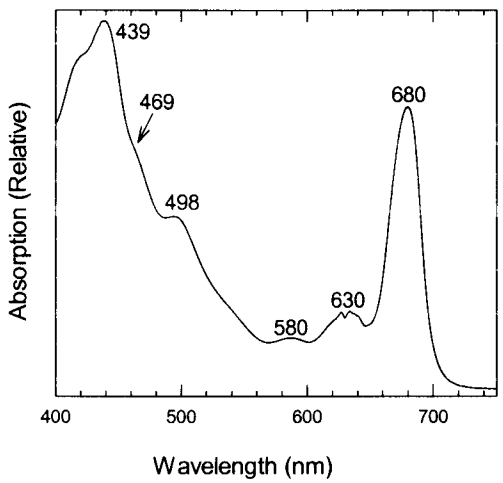

Fig 2.3. Absorption spectrum of olive brown II at room temperature.

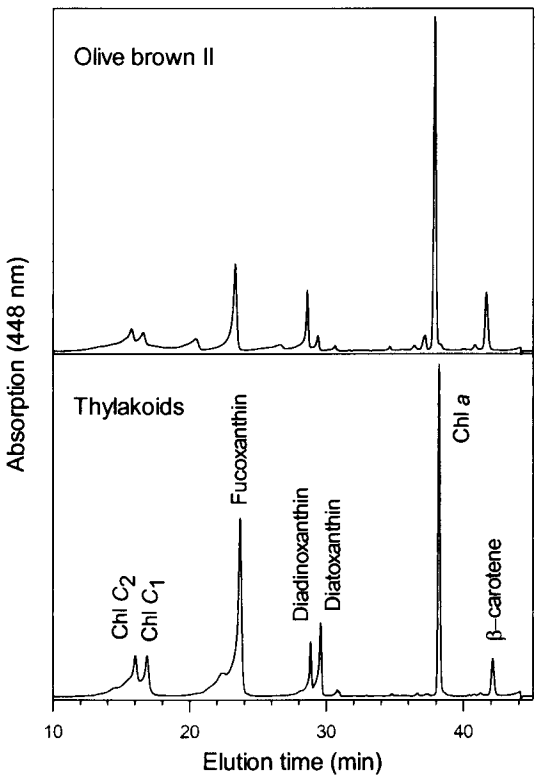

Fig. 2.4. HPLC pigment analyses of olive brown II and thylakoids. Spectra were plotted to the same relative height of Chl *a*.

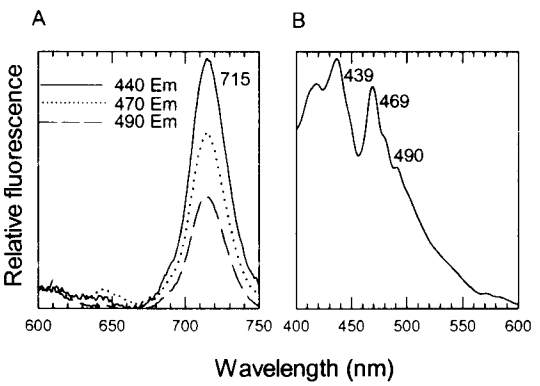

Fig. 2.5. Low-temperature (77K) fluorescence emission and excitation spectra of olive brown II. A: Emission spectra of olive brown II obtained with 440, 470, and 490 nm excitation. B: Excitation spectrum of olive brown II (715 nm emission).

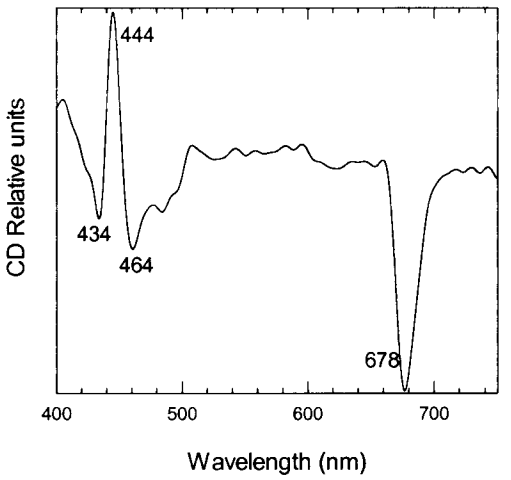

Fig. 2.6. Circular dichroism spectrum (400-750 nm) of olive brown II at room temperature.

Chapter 3 Future Prospects

In this study, a PSI-LHCI complex from the diatom *Cylindrotheca fusiformis* was isolated and characterized with respect to its spectroscopic properties and biochemical composition. PsaA, PsaB, PsaC, and nine FCPs were identified in the complex. The PSI-LHCI complex lacked detectable PSII core proteins, as determined by immunostaining and silver staining, indicating that all FCP proteins were specifically associated with PSI. Energy transfer from FCP proteins to PSI was confirmed by 77K fluorescence analysis: a single fluorescence emission peak was observed at 715 nm. This is likely from the PSI core, but could originate from an FCP with long wavelength pigments. Importantly, there were no fluorescence signals that could be associated with either free pigments (i.e., 676 and 640 nm, for Chl *a* and Chl *c*, respectively) or from dissociated FCP (i.e., at 680-683 nm). Twenty genes encoding the FCP of *Cylindrotheca* were also cloned and sequenced. The main objective of my research was to characterize *fcp* genes and proteins with an emphasis on those that specifically serve photosystem I in *Cylindrotheca.* Attempts to obtain the N-terminal sequences of FCPs that specifically associate with PSI have thus far been unsuccessful and therefore assignment of functional identities to the gene sequences is not practical at this time. Future work could include isolation and characterization of the LHCI complex, obtaining N-terminal sequences of the FCP proteins and other low molecular weight proteins in the PSI-LHCI complex, and cloning and DNA sequence analysis of the remaining *fcp* genes. Together, these future studies are needed to provide insights into the functional assignment of FCP proteins and genes as well as the structure of PSI in *Cylindrotheca fusiformis.*

Isolation and Characterization of LHCI

Information about LHCs of PSI is limited (Scheller et al. 2001), mostly because they are difficult to isolate abundantly in an intact state (Schmid et al., 1997). It is widely accepted that green plant LHCI consists of four polypeptides (Lhca 1-4) with protein

masses in the range of 20-24 kD. The cognate genes have been identified, cloned, and sequenced (Jansson, 1994). It has been shown that green plant LHCI forms two kinds of heterodimers in solution, LHCI-730 and LHCI-680. Detailed spectroscopic characterization of each component is available (Ihalaninen et al., 2000). LHCI binds PSI on only one side (Boekema et al., 2000).

LHC has been isolated and characterized from P700-enriched fractions of a diatom (Caron & Brown, 1987; Berkaloff et al, 1990). The spectral properties varied significantly from each other, and the identities of the LHCPs in these PSI-enriched fractions are debatable. The problems with prior studies relate, at least in large part, to the fact that the starting material was contaminated by PSII components. In such preparations, LHCII contamination cannot be ruled out. The PSI-LHCI (olive brown II) complex was depleted of PSII core components (CP47, CP43, D1 and D2) and the fluorescence data indicated that no uncoupled LHCII was present. Therefore, this particle should serve as a good starting material for obtaining LHCI from this diatom. Characterizing LHCI of diatoms should contribute to our understanding of LHCI structure and function. Whether diatoms have two spectroscopically different LHCI populations is not known, although a Chl *c*-containing algae seemed to have two groups of LHCI that emitted at 686 and 702 nm (Büchel and Wilhelm, 1993). Isolation of LHCI from PSI-LHCI complex should answer this question for us.

Besides the spectroscopic analysis, isolation of LHCI from PSI-LHCI complex will also contribute to our understanding of the proteins that specifically associate with PSI. Although nine proteins in PSI-LHCI cross-reacted with a LHCP antibody, we suspect that some of them serve both PSI and PSII. Further solubilization of PSI-LHCI and resolution (e.g., by sucrose density gradient centrifugation) of LHCI and the PSI core is one approach that I have tried unsuccessfully. The LHCI fraction obtained thus far appeared to be composed of unassembled FCPs and free pigments (as based on 77 K fluorescence and CD data). A slightly modified approach was also tried. Lower levels of

detergents were used, in the hopes of removing part of the LHCI complex (e.g., LHCI-680), resulting in a smaller population of LHCI associated with PSI, providing insight into which FCPs are more closely connected to the PSI core. Overall, this type of work is very important for understanding the structure and function of diatom PSI and its LHCI.

From protein to gene: the functions of FCPs

In our PSI-LHCI preparation, up to nine protein bands ranging from 27 to 10 kD cross-reacted with a *Chlamydomonas* LHCP polyclonal antibody, 20 *fcp* gene sequences (including 10 full-length cDNA sequences) have also been cloned and sequenced (see Appendix I). Of these, the deduced amino acid sequences of 8 genes match the N-terminal sequences of an 18 kD FCP and 6 match that of a 17 kD FCP (Martinson, 1996). Six other genes were found that encode unique FCPs, three of which have amino acid sequences very similar to the N-terminal sequence of a less abundant 17.5 kD FCP (Martinson, 1996). The 18 and 17 kD FCP are the most abundantly accumulated FCPs in *Cylindrotheca* and it was not too surprising that most of transcripts identified (eight of ten cDNA sequences; Appendix 1) encode one of these two proteins. There are multiple copies of the two genes (eight for the 18 kD and six for the 17 kD, see Appendix 1) in the *Cylindrotheca* genome, so another not-so-surprising result was that most of the PCR products (7 of 11; Appendix 1) obtained with FCP degenerate primers were products of these two gene families. The 18 and 17 kD FCP very likely serve both photosystems and probably serve as the primary peripheral LHC. However, no protein of 17 kD in the PSI-LHCI cross-reacted with the LHCP antibody. The reason for this could be that the 17kD protein is a PSII-specific antenna (which could also account for its abundance) that does not exist in the PSI-LHCI fraction or if it does serve PSI, was not expressed under the growth conditions employed here. There appear to be only three genes that encode 17.5 kD-like proteins, yet a protein band of 17.5 kD in the PSI-LHCI strongly cross-reacted with the LHCP antibody (Fig.2.2.c). Three genes that encode unique FCP proteins (other than the 18, 17.5 and 17 kD) have also been sequenced in *Cylindrotheca.* In addition, the

N-terminal sequence of a 22 kD protein was determined (Martinson, 1996), which does not match any protein in Genbank. These genes and proteins may represent FCPs that are more peripheral to the reaction center. Although the immuno-blotting revealed an 18 kD and a 17.5 kD FCP in the PSI-LHCI, we don't know whether they are the same proteins that Martinson (1996) sequenced. Obtaining the N-terminal sequences of the nine protein bands including the 18 and the 17.5 kD in the PSI-LHCI will help us assign functions to the genes we sequenced and the proteins that Martinson (1996) sequenced. There may be genes that exist in low copy number, or that may have very different sequences such that the primers we used did not amplify these genes. Alternatively, the transcripts could be in such low abundance that the RACE amplification was not successful. Obtaining the N-terminal sequences of these proteins should present an opportunity to identify these genes/proteins and hopefully help to determine their functions.

PSI subunits in diatoms

Eleven subunits exist in PSI of cyanobacteria and fourteen in green plants. Each of these has been characterized in terms of their function (e.g., by knock-out mutagenesis) and gene sequences. Little information is available for diatom PSI accessory subunits, let alone their functions/sequences. In fact, it is not known if the diatom PSI is more like the cyanobacterial or the green plant PSI, or if it will prove to have unique subunits not found in either of these types of organisms.

The N-terminus of a 14 kD protein has been determined from a *Cylindrotheca* PSI-enriched fraction; the sequence matched that of PsaD (Martinson, 1996). PsaD of green plants has a molecular weight of 18 kD and PsaD of cyanobacteria is 15.6 kD. The PsaD of *Cylindrotheca* is closer, in size, to that of cyanobacteria. A 9 kD protein in PSI-LHCI, PsaC, as determined in this work (Chapter 2) has the same molecular weight as PsaC of green plants and cyanobacteria. For such a highly conserved protein, with such an important function in all photosynthetic organisms (Scheller et al., 2001), it is not

surprising that both the size and N-terminal sequences are highly conserved in diatoms. Besides PsaC and PsaD, there are two additional proteins from *Cylindrotheca* PSI fractions that have been sequenced, a 10 kD from PSI-LHCI (IDRGSVV(r/v)IllEe) and an 18 kD ((Ka)F(DM)(Nds)G(LV)V(Dpg)(Fp)(Le)(GS)(YP)(Er)t(In))from PSI-enriched fraction (Martinson, 1996). The N-terminal sequences did not match any protein in the GenBank. These proteins may represent special PSI subunits from diatoms that either have unique sequences or functions. Obtaining amino acid sequence of other small molecular weight proteins observed in the PSI-LHCI should help us understand the compositions of PSI in diatoms.

As part of my project, we attempted to obtain the N-terminal amino acid sequences of several proteins present in the PSI-LHCI fraction. Besides PsaC (see above), we obtained ambiguous data for proteins in the 22-14 kD range. These poor results could be due to co-migration of PSI proteins such as PsaD, PsaF, and PsaL, which have molecular weight of 17-18 kD. Attempts to isolate LHCI after purification of the PSI-LHCI complex have thus far proven unsuccessful, but future attempts could be made to purify this antenna complex so that the amino acid sequences of its apoproteins could be determined.

References:

Berkaloff, C., Caron, L. & Rousseau, B. 1990. Subunit organization of PSI particles from brown algae and diatoms: polypeptide and pigment analysis. *Photosynth. Res.* 23:181-193.

Boekema, E. J., Jensen, P. E., Schlodder, E., Breemen, J. F. L., Roon, H., Scheller, H. V. & Dekker, J. P. 2001. Green plant photosystem I binds light-harvesting complex I on one side of the complex. *Biochemistry* 40:1029-1036.

Büchel, C. & Wilhelm, C. 1993. Isolation and characterization of a photosystem I-associated antenna (LHCI) and a photosystem I-core complex from the chlorophyll *c*-containing alga *Pleurochloris meiringensis* (Xanthophyceae). *J. Photochem. Photobiol. B Biol.* 20:87-93.

Caron, L. & Brown, J. 1987. Chlorophyll-carotenoid protein complexes from the diatom, *Phaeodactylum triconutum*: spectrophotometric, pigment and polypeptide analyses. *Plant Cell Physiol.* 28:775-785.

Ihalainen, J. A., Gobets, B., Sznee, K., Brazzoli, M., Croce, R., Bassi, R., Grondelle, R. V., Kroppi-Tommola, J. E. I. & Dekker, J. P. 2000. Evidence of two spectroscopically different dimers of light-harvesting complex I from green plants. *Biochemistry* 39:8635-8631.

Jansson, S. 1994. The light-harvesting chlorophyll *a*/*b*-binding proteins. *Biochim. Biophys. Acta* 1184:1-19.

Martinson, T. A., 1996. *Biochemistry of diatom photosynthetic membranes and pigment-protein complexes*. PhD thesis. University of Alaska Fairbanks, Fairbanks, AK.

Scheller, H. V., Jensen, P. E., Haldrup, A., Lunde, C. & Knoetzel, J. 2001. Role of subunits in eukaryotic photosystem I. *Biochim. Biophys. Acta* 1507:41-60.

Schmid, V. H. R., Cammarata, K. V., Bruns, B. U. & Schmidt, G. W. 1997. *In vitro* reconstitution of the photosystem I light-harvesting complex LHCI-730: heterodimerization is required for antenna pigment organization. *Proc. Natl. Acad. Sci. USA* 94:7667-7672.

Appendix I

fcp gene sequences of *C. fusiformis* obtained with PCR and RACE

In this appendix, 20 *fcp* gene sequences obtained with either PCR or 3'RACE are listed with their GenBank Accession numbers. The deduced amino acid sequences of these *fcp* genes encode an 18 kD FCP and a 17 kD FCP (Martinson, 1996) as well as other FCPs of *Cylindrotheca fusiformis.* The cffcpA family encoding the 18 kD protein is apparently encoded by genes at seven loci, cffcpA-1 to cffcpA-7. The cffcpB family encoding the 17 kD protein appears to be encoded by genes at six loci, cffcpB-1 to cffcpB-6. The deduced amino acid sequences of other *fcp* genes, cffcpC, cffcpD, cffcpH, are very similar to the 17.5 kD FCP (Martinson, 1996). An additional three genes encode unique FCPs that do not match the N-terminal sequences of the 18 D, the 17.5 kD, and the 17 kD FCPs.

Included in this Appendix is a diagram of a typical FCP protein/gene (Fig.A.1), the sequences of the primers used for PCR amplification (Table A.1), a brief description of each FCP cloned, including its GenBank accession number (Table A.2, A.3, A.4), multi-sequence alignments of the deduced amino acid sequences with the N-terminal sequences of the 18 kD and the 17 kD FCP (Fig A.2), and the putative AAUAAA cleavage and polyadenylation specificity factor (CPSF) present in the 3' UTR of *fcp* RACE products (Fig A.3).

Figure A.1: Typical *fcp* gene and protein structure (upper) and PCR primer sites of PCR and RACE products (lower). TMH1-3: Trans-membrane Helix 1-3 (upper); arrows indicated primer sites (lower) as described in more detail in Table A1.

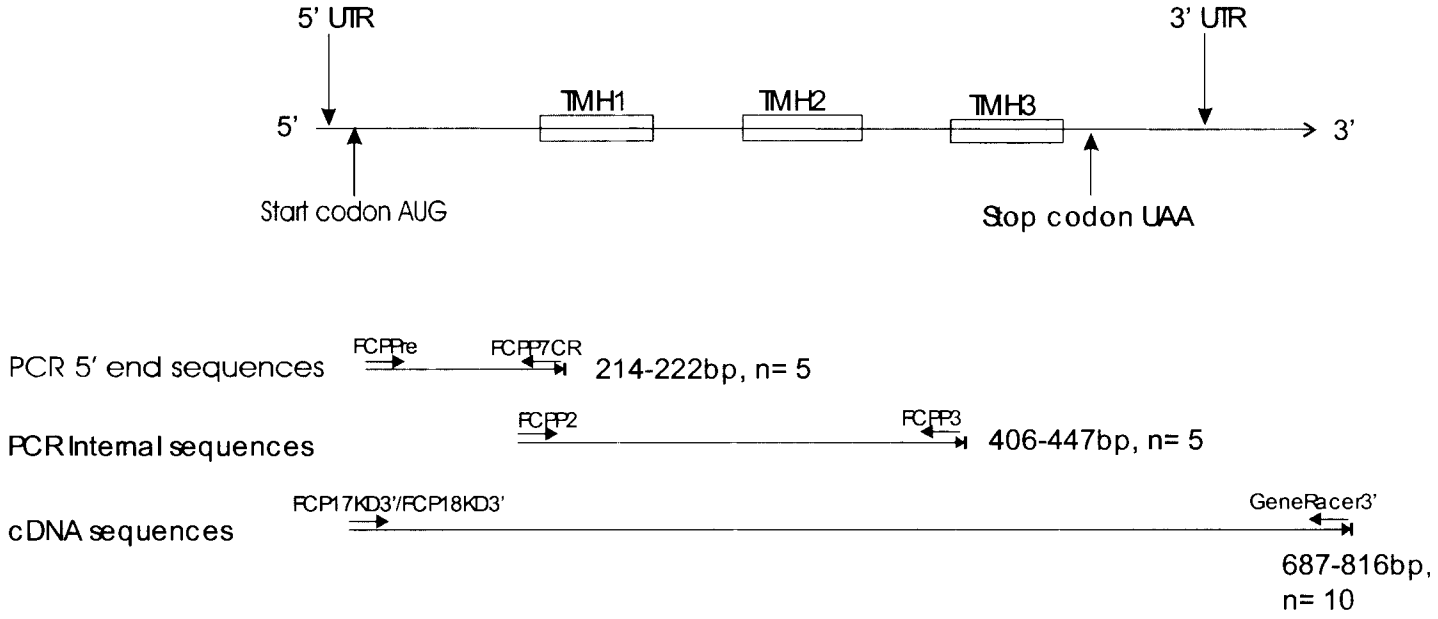

Table A.1: Primers used to generate PCR and RACE products.

Primer Region	Primer Name	Primer Sequence*
5' end Forward	FCPPre (21mer)	5'-ATGAARHYYRHYRTBHTYGCC-3'
5' end Reverse	FCPP7CR (20mer)	5'-MSMGATRCGTCCGTGCTTGA-3'
Internal Forward	FCPP2 (21mer)	5'-CCNYTNGGNTTYTTYGAYCC-3'
Internal Reverse	FCPP3 (21mer)	5'-CATNARNGCNARDATNCCCAT-3'
3' RACE Forward	FCP18KD3' (28mer)	5'-ATGAAGWCYGWYGTRCTYGCCTCCCTTG-3'
3' RACE Forward	FCP17KD3' (28mer)	5'-ATGAAGMYYRHYGTBHTYGCCKCCCTTA-3'
3' RACE Reverse	GeneRacer 3' primer (25mer)	5'-GCTGTCAACGATACGCTACGTAACG-3'

* B: A/G; D: A/G/T; H: A/C/T; M:A/C; R:A/G; S:C/G; Y:C/T; W: A/T; N: A/C/T/G;
GeneRacer 3' primer is from Novagen.

Table A.2: Sequences encoding the 18 kD FCP (Martinson, 1996) of *Cylindrotheca fusiformis*. The N-terminal sequence of the 18 kD FCP: NAFESELGAQPPLGFFDPFGLLSGDATEER. For BLAST, the entire deduced amino acid sequence was used for protein searches while the entire DNA sequence, minus degenerate primer regions, was used for DNA searches.

Fragment Name	Gene Name	Length (bp)	Coding region (CDS)	Fragment Type	GenBank Accession Number	Similarities (at DNA Level)	Similarities (at Protein Level)
3R18KD-3	cffcpA-1A	816	1-629 (208 AA)	3'RACE	AY125580	Skeletonema costatum scfcpH, U66180	Cyclotella cryptica CCFCP5, CAA04401
3R18KD-4	cffcpA-2A	805	1-629 (208 AA)	3'RACE	AY125581	Skeletonema costatum, scfcpF, U66178	Same as above
3R18KD-2	cffcpA-2	760	1-633 (210 AA)	3'RACE	AY125582	Thalassiosira pseudonana 3HfcpB, U66184	Same as above
3R18KD-5	cffcpA-3	794	1-633 (210AA)	3'RACE	AY125583	Skeletonema costatum scfcpH, U66180	Same as above
Z-M	cffcpA-4	222	1-222 (74 AA)	PCR, 5'end	AY125584	Skeletonema costatum scfcpG, U66179	Same as above
Z-O	cffcpA-5	222	1-222(74 AA)	PCR, 5'end	AY125585	Same as above	Same as above
Z-7	cffcpA-6	214	1-214 (71 AA)	PCR, 5'end	AY125586	Same as above	Same as above
H-2	cffcpA-7	447	1-447 (149 AA)	PCR, internal	AY125587	Skeletonema costatum scfcpF, U66178	Same as above

Table A.3: Sequences encoding the 17 kD FCP (Martinson, 1996) of *Cylindrotheca fusiformis*. The N-terminal sequence of the 17 kD FCP: SFENELGAQPPLGFFDPLGLVADGDQEKFDRL. For BLAST, the entire deduced amino acid sequence was used for protein searches while the entire DNA sequence, minus degenerate primer regions, was used for DNA searches.

Fragment Name	Gene Name	Length (bp)	Coding regions (CDS)	Fragment type	GenBank Accession number	Similarities (at DNA Level)	Similarities (at Protein Level)
3R17KD-3	cffcpB-1	741	1-594 (197AA)	3'RACE	AY125588	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Phaeodactylum triconutum FCPF, Q41094
3R17KD-5	cffcpB-2	752	1-594 (197 AA)	3'RACE	AY125589	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Phaeodactylum triconutum FCPF, Q41094
3R18KD-1	cffcpB-3	741	167 or 197 AA	3'RACE	AY125590	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Cyclotella cryptica Ccfcp3.CAA0440
3R18KD-12	cffcpB-4	748	1-594 (197 AA)	3'RACE	AY125591	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Phaeodactylum triconutum FCPF, Q41094
Z-E	Partial cffcpB-1	432	1-432 (144 AA)	PCR, internal	N/A	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Phaeodactylum triconutum FCPF, Q41094
H-1	cffcpB-5	408	1-408 (135 AA)	PCR, internal	AY125592	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Phaeodactylum triconutum FCPF, Q41094
Z-4	cffcpB-6	216	1-216 (72 AA)	PCR, 5'end	AY125593	Phaeodactylum triconutum fcpE and fcpF genes Z23153	Phaeodactylum triconutum FCPE, Q41093

Table A.4: Sequences encoding other FCPs of *Cylindrotheca fusiformis*. For BLAST, the entire deduced amino acid sequence was used for protein searches while the entire DNA sequence, minus degenerate primer regions, was used for DNA searches.

Fragment name	Gene name	Length (bp)	Coding regions (CDS)	Fragment type	GenBank Accession number	Similarities (at DNA Level)	Similarities (at Protein Level)
3R17KD-2	cffcpC	687	1-583 (194AA)	3'RACE	AY125594	Thalassiosira pseudonana, 3HfcpA, U66183	*Cyclotella cryptica* CAA04178
3R18kD-15	cffcpD	734	1-591 (197 AA)	3'RACE	AY125595	Phaeodactylum triconutum fcpE and fcpF genes, Z23153	*Phaeodactylum triconutum* FCPF, Q41094
Z-17/Z-20	cffcpE	414	1-414 (138AA)	PCR, internal	AY125596	Petunia hybrida pCab10 K00975	Nannochloropsts sp. Violaxanthir/Chl a binding protein, AAB94637
Z-2/Z-14	cffcpF	207 /216	1-207 (69AA) /1-216 (72 AA)	PCR, 5'end	AY125597	Skeletonema costatum ScfcpB, U66171	*Skeletonema costatum*, ScfcpA, AAB80924
Z-B	cffcpG	424	1-424 140AA	PCR, internal	AY125598	*Skeletonema costatum* scfcpG, U66179	*Odontella sinensis* Q42395
Z-19	cffcpH	433	1-433 (144 AA)	PCR, internal	AY125599	Odontella sinensis fcpB gene, X81055	*Cyclotella cryptica* CAA04178

Fig. A.2. Multiple protein sequence alignment of PCR and RACE products with the N-Terminal sequences of the 17 kD and 18 kD FCPs.

```
17kD-N ter:  ----------------------------SFENELGAQPPLGFFDPLGLVA-DGDQEKFDRL---------------------------------------- :  32
cffcpB-1:    MKTIVLAALIATAAAFAPAKQAATTTSLN-SFENELGAQPPLGFFDPLGLVA-DGDQEKFDRLRYVELKHGRICMLGVVGYLVTAAGIRLPGDIDYSGTAFSDIT : 103
cffcpB-2:    MKIDVIASLIATAAAFAPAKQAATTTSLN-SFENELGAQPPLGFFDPLGLVA-DGDQEKFDRLRYVELKHGRICMLGVVGYLVTAAGIRFPGDIDYSGTAFSDIT : 103
cffcpB-3:    MKSDVLASLVATAAVFAPAKQAATTTSLN-SFENELGAQPPLGFFDPLGLVA-DGDQEKFDRLRYVELKHGRICMLGVVGYLVTAAGIRLPGDIDYSGTAFSDIT : 103
cffcpB-4:    MKSDVLASLVATAAAFAPAKQAATTTSLN-SFENELGAQPPLGFFDPLGLVA-DGDQEKFDRLRYVELKHGRICMLGVVGYLVTAPGIRLPGDIDYSGTAFSDIT : 103
cffcpB-6:    MKPNVIASLIATAAAFAPAKQAATTTSLN-SFENELGAQPPLGFFDPLGLVA-DGDQEKFDRLRYVELKHGRIA---------------------------- :  72
cffcpB-5:    ------------------------------------------FFDPLGLVA-DGDQEKFDRLRYVELKHGRICTLGVVGYLVTAAGIRLPGDIDYSGTAFSDIT :  61
18kD-N ter:  ----------------------------NAFESELGAQPPLGFFDPFGLLSGDATEER---------------------------------------- :  30
cffcpA-1A:   --SVVLASLVATAAAFAPAKQASTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGRICMLAFLGQITTRAGIHLPGSINYAGDSFDSYP : 103
cffcpA-1B:   --SDVLASLVATAAAFAPAKQASTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGRICMLAFLGQITTRAGIHLPGSINYAGDSFDSYP : 103
cffcpA-2:    MKTDVLASLVATAAAFAPAKQASTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGRICMLAFLGQITTRAGIHLPGSINYAGDSFDSYP : 105
cffcpA-3:    MKTDVLASLVATAAAFAPAKQASTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGRICMLAFLGQITTRAGIHLPGSINYAGDSFDSYP : 105
cffcpA-5:    MKSDVLASLVATAAAFASAKQATTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGRIA---------------------------- :  74
cffcpA-4:    MKTDVLASLVATAAAFAPAKQASTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGRIS---------------------------- :  74
cffcpA-6:    ---VVFASLVATAAAFAPAKQATTSTALNNAFESELGAQPPLGFFDPFGLLSGDATEERFERLRYVEIKHGR------------------------------ :  69
cffcpA-7:    ----------------------------------------PLGFFDPFGLLSGDATEERFERLRYVEIKHGRICMLAFLGQITTRAGIHLPGSINYAGDSFDSYP :  65
cffcpH:      ----------------------------------------PLGFFDPLGLVENA-DQAKFDRLRFVELKHGRICMLGVVGYLVTAAGIRLPGDIDYSGTAFSDIT :  64
cffcpD:      MKSDVLASLVATAAAFAPAKQAATTTSLN-SFENELGAQPPLGFFDPLGPVADG-DQEKFDRLRYVELKHGRICMLGVVGYLVTAAGIRLPGDIDYSGTAFSDIT : 103
cffcpC:      -----FAALIATAAAFAPAKQAATTTSLN-SFESELGAQDPLGFFDPLGLVENA-DQAKFDRLRFVELKHGRICMLGVLGYLVTAAGVRLPGDIDYAGTKFADIG :  98
cffcpG:      -------------------------------------------------LGFLADA-DQERFDRLRYVELKHGRIAMPAFLGQITTRAGIHLPGNIDLAGDSFDSYP :  57
cffcpF:      MKLVVFAALISAAAAFAPAKQAAQTTSL-SAFESELGAQEPLGFYDPLGFLADA-DQERFDRLRYVELKHGRIA---------------------------- :  72
cffcpE:      ----------------------------------------PLGFFDLLGFSKNKSDETM-QHYRESELKHGRVAMAACLGWYLNAAGVHPAFNSELSNNPLEAAQ :  64
```

```
17kD-N ter:  ----------------------------------------------------------------------------------------------------
cffcpB-1:    YGWE---GSQQVPVAGALQVFAFVGFLELAVMKDITG-GEFPGDFRNGALDFGWDSFDEETKLQKRAVELNQGRAAQMGMLALMVHDKLG--NVENFFPN----- : 197
cffcpB-2:    YGWE---GSQQVPVAGALQVFAFVGFLELAVMKDITG-GEFPGDFRNGALDFGWDSFDEETKLQKRAVELNQGRAAQMGMLALMVHDKLG--NVENFFPN----- : 197
cffcpB-3:    YGWE---GSQQVPVAGALQVFAFVGFLELAVMKDITG-GEFPGDFRNGALDFGWDSFDEETKLQKRAV-LNQGRAAQMGMLALMVHDKLG--NVENFFPN----- : 196
cffcpB-4:    YGWE---GSQQVPVAGALQVFAFVGFLELAVMKDITG-GEFPGDFRNGALDFGWDSFDEETKLQKRAVELNQGRAAQMGMLALMVHDKLG--KVENFFPN----- : 197
cffcpB-6:    ---------------------------------------------------------------------------------------------------- :   -
cffcpB-5:    YGWE---GSQQVPVAGALQVFAFVGFLELAVMKDITG-GEFPGDIRNGALDFGWDSFDEETKLQKRAVELNQGRAAQM---------------------- : 135
18kD-N ter:  ---------------------------------------------------------------------------------------------------- :   -
cffcpA-1A:   NGVAAVFGENHIPTAGIVQIISTIGILECAFMREVPGFAEFPGDFRNGYIDFGWDDFDEETKLQKRAIELNNGRAAMMGILGLMVHELIIPLGYDPALPIIGHLQ : 208
CFFCPa-1B:   NGVAAVFGENHIPTAGIVQIISTIGILECAFMREVPGFAEFPGDFRNGYIDFGWDDFDEETKLQKRAIELNNGRAAMMGILGLMVHELIIPLGYDPALPIIGHLQ : 208
cffcpA-2:    NGVAAVFGENHIPTAGIVQIISTIGILECAFMREVPGFAEFPGDFRNGYIDFGWDDFDEETKLQKRAIELNNGRAAMMGILGLMVHELIIPLGYDPALPIIGHLQ : 210
cffcpA-3:    NGVAAIFGDNHIPTAGIVQIISTIGILECAFMREVPGFAEFPGDFRNGYIDFGWDDFDEETKLQKRAIELNNGRAAMMGILGLMVHELIIPLGYDPELPIIGHLQ : 210
cffcpA-5:    ---------------------------------------------------------------------------------------------------- :   -
cffcpA-4:    ---------------------------------------------------------------------------------------------------- :   -
cffcpA-6:    ---------------------------------------------------------------------------------------------------- :   -
cffcpA-7:    NGVAAVFGENHIPTAGIVQIISTIGILECAFMREVPGFAEFPGDFRNGYIDFGWDDFDEETKLQKRAIELNNGRAAMMGILALM---------------------- : 149
cffcpH:      YGWE---GSQQVPVAGALQVFAFVGFLELAVMKD-ITGGEFPGDFRNGALDFGWDSFDEETKLQKRAVELNQGRAAQMGILALM---------------------- : 144
cffcpD:      YGWE---GSQQVPVAGALQVFAFVGFLELAVMKD-ITGGEFPGDFRNGALDFGWDSFDEETKLQKRAVELNQGRAAQMGMLALMVHDKLG--NVENFFPN----- : 197
cffcpC:      YGWE---GSQQVPVAGALQIFAFCGFLELGVMKDVKGTGEFIGDFRNGALDFGWDKFDAETKLQKRAIELNNGRAAQMGMLGLMIHDKLG--NVESLFPGN---- : 194
cffcpG:      DGLAALFGEDAIPSAGLAQIVGFIGCLELAVMKDVKGEGEFPGDFRNGAIDFGWDTFGPETKLRKRAIELNNGRAAMMGILALM---------------------- : 141
cffcpF:      ---------------------------------------------------------------------------------------------------- :   -
cffcpE:      ----------QLPAVGWLQFVLGCGAIEWLAEKIKERPGYKAGDFLGAAYWTDDSDELWVGYQNREINNGRLAMVAFMGILALM---------------------- : 138
```

Figure A.3: Putative AAUAAA cleavage and polyadenylation specificity factor (CPSF) in 3' UTR of *fcp* RACE products.

```
                    *         20         *         40         *         60         *         80         *        100
3R17KD-3:  : TAAGCTGCTACACCATCCGAACAGGACAAGCCGCGTTGTCCGATAGGTTTGGAATGGAGGAGGTTAGACCTACCGACTCCCCCATTTGCTTGCTCTAGGT--------- : 100
3r18-12:   : TAAGCTGCTACACCATCCGAACAGGACAAGCCGCGTTGTCCGATAGGTTTTGAATGGAGGAGGTTAGACCTACCGACTCCCCCATTTGCTTGCTCTAGGT--------- : 100
3r18kd-1:  : TAAGCTGCTACACCATCCTAACAGGACAAGCCGCGTTGTCCGATAGGTTTTGAATGGAGGGGGTTAGACCCACAGACTCTCCCATTTGCTTGCTCTAGGT--------- : 100
3r17kd-5:  : TAAGGTGATACACCATCCTAACAGGACAAGCCGCGTTGTCCGATAGGTTTTGAATGGAGGGGGTTAGACCCACAGACTCTCCCATTTGCTTGCTCTAGGT--------- : 100
3r18-3:    : TAAATTGCTTGTTGAGAAATAGGAAAAACTGTGGCCCTCTTTGAATGTAAACAGTTCGAATGATGCGGACCTTTTTTTCCTACCACGCAAATGCAACCAGGTCGACCAA : 109
3r18-4:    : TAAATTGCTTGTTGAGAAATAGGAAAAACTGTGGCCCTCTTTGAATGTAAACAGTTCGAATAATGCGGACCTTTTTTTCCTACCACGCAAATGCAACCAGGTCGACCAG : 109
3r18-5:    : TAAATTGCTTGCAAGTGCGAGATGGCCGTTGGTTGAATGATGGATTTCAAAAGACTAGGGCTACCAGTCTTCCGTTTTTCTCGCTTAGTA------------------- :  90
3r18-2:    : TAAATTGACTATTTGTCTTTATCCGCTGAAAGACTGCAGTTCAATCGGCTACTACTAGACACCCTAAGATGATTCGTCCCAAAGAGCGCT------------------- :  90
3r17kd-2:  : TAAGCGTTCGAAGCTAAAGATTGGCCCGTTGTAGATCGTTTTTTACACCGATTT------------------------------------------------------- :  54
3r18-15:   : TAAGCTGCTACAACAATCTAGGAGAAACCGCGTTGTCCGACGAGTTTTGAGCACCGATGTCGGCCGCATGTTGTCTGCTTC---------------------------- :  81
Consensus    TAA  tg t
```

```
            *       120         *       140         *       160         *       180         *       200         *
3R17KD-3:  : ----------------AATTCTACTAATAAGTATTGAGTTATCTTGGCTG---------AAAAAAAAAAAAAAAA---------------------------- : 150
3r18-12:   : ----------------AATTCTACTAATAAGTATTGAGTTATCTTGGCTG---------AAAAAAAAAAAAAAAAAAAAAAA------------------- : 157
3r18kd-1:  : ----------------AATTCTACTAATAAGTATTGAGTTATCTTGGTT----------AAAAAAAAAAAAAAAAA--------------------------- : 150
3r17kd-5:  : ----------------AATTCTAGTAATAAGTATTGAGTTATCTTGGTT----------AAAAAAAAAAAAAAAAAAAAAAAAAAAAA--------------- : 161
3r18-3:    : CCTTCCGTTTTCTTGCGTAGTCTTTAAATATATTAATAGATGAAGAGAGTTACATGG---AAAAAAAAAAAAAAAAAAAAAAAAA------------------- : 190
3r18-4:    : CCTTCCGTTTTCTTGCGTAGTCTTTAAATAAATTAATAGATGAAGAGAGTT---------AAAAAAAAAAAAAAAAAAA------------------------ : 179
3r18-5:    : ------------------TAAAAGAAATAGATGAAGAGAGTTTCAACGG----------AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA : 164
3r18-2:    : --------------------------AATAGGACTTGAGTTATCC---------------AAAAAAAAAAAAAAAAAAAAA---------------------- : 130
3r17kd-2:  : --------------------------AATAGTACTACTCTAGTTCAAAGTTACCTTGGTTAAAAAAAAAAAAAAAAA--------------------------- : 104
3r18-15:   : ------------------AAACTAAATCTACTAGAGTATTGAGTTAGTTTGC------AAAAAAAAAAAAAAAAAAAAAAAAAAAAA--------------- : 143
Consensus                        AATa       ag t       t         AAAAAAAAAAAAAAAAA
```